National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Services des thèses canadiennes

Ottawa, Canada
K1A 0N4

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

Canada

NL-339 (r.86/06)

THE UNIVERSITY OF ALBERTA

Studies of Transfer (tra) Genes from IncF Plasmids

by

B. Brett Finlay

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

EDMONTON, ALBERTA
FALL, 1986

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-32470-8

University of Alberta
Edmonton
Canada T6G 2H7

Department of Biochemistry
Brett Finlay
474 Medical Sciences Building, Telephone (403) -432-4574

Feb. 26, 1986

RECEIVED
MAR 5 1986
PUBLICATIONS

Director of Publications
American Society for Microbiology
1913 I St., NW
Washington, D.C. 20006

Dear Sir/Madam:

I am currently preparing my doctoral thesis in which I wish to include material previously published in the Journal of Bacteriology. To do so, I require a letter of copyright permission for the following manuscripts:

1. Finlay, B., Paranchych, W., Falkow, S. 1983. Characterization of conjugative plasmid EDP208. J. Bacteriol. 156: 230-235.
2. Finlay, B., Frost, L., Paranchych, W. 1984. Localization, cloning and sequencing of the conjugative plasmid ColB2 pilin gene. J. Bacteriol. 160:402-407.
3. Finlay, B., Frost, L., Paranchych, W., Parker, J., Hodges, R. 1985. The major antigenic determinants of F and ColB2 pili. J. Bacteriol. 163:331-335.
4. Finlay, B., Frost, L., Paranchych, W. 1986. Nucleotide sequence of the R1-19 plasmid transfer genes traM, finP, traJ, traY and traYZ promoter. J. Bacteriol. May issue.

The manuscripts will be included in a form similar to the published version and will be accompanied by a footnote stating that "a version of this chapter has been published elsewhere". This will be followed by the appropriate reference.

Thank you for your prompt attention.

Sincerely,

Brett Finlay

Brett Finlay

A specific citation to the journal must be made.

PERMISSION GRANTED CONTINGENT ON AUTHOR PERMISSION
American Society for Microbiology
Publications Department

Director _________ Date 3-10-86

University of Alberta
Edmonton

Canada T6G 2H7

Department of Biochemistry
B. Finlay

474 Medical Sciences Building, Telephone ~~(403)~~
(403)-432-4574

RECEIVED
APR 14 1986
PUBLICATIONS

April 6, 1986

Director of Publications
American Society for Microbiology
1913 I St., NW
Washington, D.C. 20006

Dear Sir/Madam:

I am currently preparing my doctoral thesis in which I wish to include material previously published in the Journal of Bacteriology. To do so, I require a letter of copyright permission for the following manuscript:

Finlay, B., and W. Paranchych. 1986. Nucleotide sequence of the surface exclusion genes traS and traT from the IncF$_{O}$ lac plasmid pED208. J. Bacteriol. June issue.

The manuscript will be included in a form similar to the published version and will be accompanied by a footnote stating that "a version of this chapter has been published". This will be followed by the appropriate reference.

Thank you for your prompt attention.

Sincerely,

Brett Finlay

Brett Finlay

PERMISSION GRANTED CONTINGENT ON AUTHOR PERMISSION
American Society for Microbiology
Publications Department

Director ____________ Date 4/16/86

CO-AUTHOR PERMISSION FORM

I, the undersigned, hereby grant permission to B. Brett Finlay to use material from:

1. Finlay, B., Frost, L., Paranchych, W., Parker, J., Hodges, R. 1985. The major antigenic determinants of F and ColB2 pili. J. Bacteriol. 163:331-335.

in his doctoral thesis.

Signed ......Robert S Hodges......

Name ......Robert S. Hodges......

Date ......March 4/86......

CO-AUTHOR PERMISSION FORM

I, the undersigned, hereby grant permission to B. Brett Finlay to use material from:

1. Finlay, B., Paranchych, W., Falkow, S. 1983. Characterization of conjugative plasmid EDP208. J. Bacteriol. 156: 230-235.

in his doctoral thesis.

Signed Stanley Falkow

Name Stanley Falkow

Date 3/5/86

CO-AUTHOR PERMISSION FORM

I, the undersigned, hereby grant permission to B. Brett Finlay to use material from:

1. Finlay, B., Frost, L., Paranchych, W., Parker, J., Hodges, R. 1985. The major antigenic determinants of F and ColB2 pili. J. Bacteriol. 163:331-335.

in his doctoral thesis.

Signed ....

Name ....J. M. R. PARKER....

Date ....FEB 26th/86....

CO-AUTHOR PERMISSION FORM

I, the undersigned, hereby grant permission to B. Brett Finlay to use material from:

1. Finlay, B., Frost, L., Paranchych, W. 1984. Localization, cloning and sequencing of the conjugative plasmid ColB2 pilin gene. J. Bacteriol. 160:402-407.
2. Finlay, B., Frost, L., Paranchych, W., Parker, J., Hodges, R. 1985. The major antigenic determinants of F and ColB2 pili. J. Bacteriol. 163:331-335.
3. Finlay, B., Frost, L., Paranchych, W. 1986. Nucleotide sequence of the R1-19 plasmid transfer genes traM, finP, traJ, traY and traYZ promoter. J. Bacteriol. May issue.

in his doctoral thesis.

Signed ......Laura Frost......................

Name ......LAURA FROST......................

Date ......February 26, 1986..........

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR:	B. Brett Finlay
TITLE OF THESIS:	Studies on Transfer (tra) Genes from IncF Plasmids
DEGREE:	Doctor of Philosophy
YEAR THIS DEGREE GRANTED	Fall, 1986

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) Brett Finlay...

PERMANENT ADDRESS:
c/o Dept. of Biochemistry
University of Alberta
Edmonton, ALberta
T6G 2H7

DATE: May 10, 1986

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Facutly of Graduate Studies and Research, for acceptance, a thesis entitled Studies on Transfer (*tra*) Genes from IncF Plasmids submitted by B. Brett Finlay in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Wm Paranchych

Supervisor

A.R. Morgan

Diane E. Taylor

Reinhart Reithmeier

External Examiner

Date: May 8, 1986

For my wife, Jane

and

my parents, Joy and Cam

ABSTRACT

Conjugative plasmids are large molecules of extrachromosomal DNA which encode the ability to transfer a copy of its DNA to a recipient bacterium. Besides the genes necessary for conjugation (the transfer (_tra_) region), genes encoding antibiotic resisitances, toxins, iron metabolism and various virulence factors are often contained on these plasmids. The object of this thesis was to clone and characterize some of the genes necessary for the bacterial conjugation process.

One of the conjugative plasmids used for this study was pED208. A restriction map of this IncFV plasmid was generated and transposon mutagenesis was employed to determine the size and location of the region of this plasmid required for pilus production. Four large _Hin_dIII fragments, which comprise 93% of pED208 plasmid DNA, were cloned and characterized. This allowed localization of the _lac_, _oriV_, _oriT_ and _tra_ genes. _lac_ and _oriV_ (the origin of vegetative plasmid replication) were encoded by the same _Hin_dIII fragment, while _oriT_ (the origin of plasmid transfer) and _tra_ were located on separate _Hin_dIII fragments.

The pED208 _tra_ region was subcloned further and the surface exclusion locus identified. This region encodes two surface exclusion genes, _traS_ and _traT_ which were sequenced. The gene product of pED208 _traT_ is produced in

large quantities, and is localized in the cell envelope. This product shares 80% homology with the traT products from the IncF plasmids F and R-100. However, these products provide low levels of exclusion cross-reactivity, indicating that the few differences occurring between these predicted gene products are sufficient to provide different traT alleles. Secondary structure predictions of the traS gene product suggested that it is a membrane protein. This protein was not detectable in minicell experiments, although chimeras containing this gene exhibit exclusion properties.

The region encoding the pED208 pilin gene (traA) was cloned and sequenced. The predicted traA product encodes a 55 residue leader peptide and a 64 amino acid pilin protein. This pilin protein shares some homology with other IncF pilin sequences. The traL and traE genes were also sequenced and their predicted gene products are 44% homologous to the equivalent products encoded by the F plasmid. pED208 traY contained an IS2 insertion element in it, producing a predicted IS2-traY fusion product which presumably provides the observed traYZ endonuclease activity. This IS2 element is in orientation II, and contains a promoter that appears to cause transcription of the pED208 tra operon. This transcription uncouples the tra operon from finOP/traJ operon control, and produces a derepressed state which results in extensive piliation of cells containing the pED208 plasmid.

A restriction map of the IncFII plasmid ColB2 was generated and the ColB2 _traA_ gene was cloned and sequenced. ColB2 pilin contains an acetylated amino terminus and is highly homologous to the F _traA_ product. The only differences in these proteins occur at the N-terminus, with the ColB2 sequence being N-Ac-ala-gln-gly-gln- while the F sequence is N-Ac-ala-gly-ser-ser-gly-gln-.

The differences found at the N-termini of the F and ColB2 pilin molecules account for the serological differences reported for these proteins. Natural and synthetic peptides corresponding to the N-terminal F and ColB2 pilin sequences were constructed and used in competition immunosorbant assays, proving that the N-terminal acetyl group and first few amino-terminal residues of F and ColB2 pilin constitute the major antigenic determinants of these proteins.

Several other genes of IncF plasmids involved in regulation of conjugation and conjugal DNA transfer were cloned and sequenced. These genes include _traM_, _traJ_, _traY_, and _traYZ_ promoter from R1-19, _oriT_, _traY_ and _traM_ from ColB4, _traY_ from R-100 and _finP_s from R124, ColVB_trp_, R1-19, R-100, ColB2 and ColB4. All of the predicted _traM_ products were approximately 80% homologous to other IncF _traM_ proteins, with the differences defining _traM_ specificity being localized in the amino terminal domain. There was more variation in the predicted _traY_ products, although several residues were conserved. Comparison of

IncF *traM* and *traY* sequences with *oriT* sequences identified potential *traM* and *traY* product binding sites. IncF *finP* products appear to be anti-sense RNA molecules, being complementary to part of the *traJ* gene. *finP* specificity is determined by a variable 6 to 7 nucleotide loop that is part of a stem-and-loop structure of the *finP* RNA molecule. Sequence information of FinP$^-$ mutations supported this concept and also identified the *finP* promoter. There are three *traJ* and *traYZ* promoter alleles within IncF plasmids. These alleles differ significantly in their sequences, although they appear to operate by analogous mechanisms.

ACKNOWLEDGMENTS

I would like to thank Dr. W. Paranchych for providing me with the opportunity to carry out this research, and also for his continued guidance and encouragement. I would also like to thank Dr. L. Frost for her valuable scientific comments, as well as her patience. Thanks are due to my fellow graduate students Dr. G. Armstrong, Dr. T. Watts, Dr. E. Worobec, B. Pasloske, and Q. Sun for contributing socially and scientifically to my graduate career. I am grateful to K. Volpel for her excellent technical assistance.

I would also like to extend my gratitude to the following people who assisted in this work: R. Herbert, for his assistance with the Tn5 mutations; A. Opgenorth, for initiating the R1-19 sequencing studies; Drs. D. Portnoy and S. Falkow for their advice on molecular genetic techniques; Drs. R. Hodges and R. Parker for providing synthetic peptides; M. Nattriss for amino acid analyses; and Drs. K. Ippen-Ihler, N. Willetts, E. Minkley and W. Dempsey for their valuable discussions on conjugation.

Finally, I would like to thank my family for their continual support and encouragement throughout my graduate training, and especially my wife, Jane.

This work was supported by the Alberta Heritage Foundation for Medical Research in the form of a studentship and a research allowance.

TABLE OF CONTENTS

Page

Abstract v
Acknowledgements ix
Table of Contents. x
List of Figures. xvii
List of Tables xx

Chapter I. Introduction. 1
    A. Pilus biosynthesis. 3
    B. Mating pair stabilization 5
    C. Conjugal DNA metabolism 6
    D. Surface exclusion 9
    E. Regulation of transfer. 11
    F. Bibliography. 16

Chapter II. Characterization of conjugative plasmid pED208 24
    A. Introduction. 24
    B. Materials and Methods 25
        1. Bacterial strains. 25
        2. Isolation of plasmid DNA 25
        3. Restriction endonuclease digests 26
        4. Fragment Isolation 26
        5. Recombinant DNA techniques 27

        - a. Cloning HindIII fragments (pBF101-105) 27
        - b. Cloning HindIII/XhoI fragment (pBF106) 27
        - c. Cloning XhoI fragment (pBF111) . 28
    - 6. f1 phage sensitivity tests 28
    - 7. pED208::Tn1 mutagenesis. 28
    - 8. pED208::Tn5 mutagenesis. 29
    - 9. Mating tests 29
- C. Results and Discussion. 30
    - 1. Restriction enzyme analysis of pED208. . 30
    - 2. Transposon mutagenesis to identify the pilus-producing region 32
    - 3. Location of Lac, oriV, and pilus-producing regions of the plasmid 33
    - 4. Complementation of adjacent HindIII fragments resulting in transfer. 35
- D. Bibliography. 45

Chapter III. Localization, cloning, and sequence determination of the conjugative plasmid ColB2 pilin gene. 48

- A. Introduction. 48
- B. Materials and Methods 49
    - 1. Bacterial strains and plasmids 49
    - 2. Isolation of plasmid DNA 49
    - 3. Recombinant DNA techniques 50

4. DNA filter hybridization 50

5. Colony filter hybridization. 50

6. Chimeras using pUC8. 50

7. M13 Dideoxy sequencing 51

8. ColB2 pilin purification 51

9. Enzymatic digestion of pilin 51

10. Tryptic peptide Sequencing. 51

11. Complementation assay 52

C. Results 52

1. Restriction enzyme analysis of ColB2 . . 52

2. DNA homology with the F tra region . . . 53

3. Cloning of traA of ColB2 53

4. Sequencing of the ColB2 pilin gene . . . 58

5. Sequence of tryptic peptides from ColB2 pilin. 58

6. The N and C- terminals of ColB2 pilin. . 59

D. Discussion. 61

E. Bibliography. 74

Chapter IV. Major antigenic determinants of F and ColB2 pili 78

A. Introduction. 78

B. Materials and methods 80

1. Bacteria 80

2. Preparation of whole pili. 81

3. Antisera production. 81

4. Antisera preabsorption 81

5. Competition ELISA. 82
6. Amino terminal tryptic peptide isolation 82
7. Cyanogen bromide digestion 82
8. Peptide synthesis and purification . . . 82
9. Amino acid analysis. 83
C. Results 84
D. Discussion. 90
E. Bibliography. 102

Chapter V. Nucleotide sequence of the R1-19 plasmid transfer genes traM, finP, traJ, traY, and traYZ promoter 106
A. Introduction. 106
B. Materials and methods 109
1. Bacterial strains and plasmids 109
2. Plasmid purification 109
3. Recombinant DNA techniques and DNA sequencing 110
4. Galactokinase (galK) assay 110
C. Results .110
1. Cloning and sequencing strategies. . . .110
2. R1-19 traM gene. 113
3. R1-19 finP gene. 114
4. R1-19 traJ gene. 115
5. R1-19 traYZ promomter. 117
6. R1-19 traY gene. 119
D. Discussion. 120

E. Bibliography. 133

Chapter VI. Nucleotide sequences of five IncF plasmid *finP* alleles 139
A. Introduction. 139
B. Results and discussion. 140
C. Bibliography. 151

Chapter VII. Studies on the origin of transfer of IncF plasmids: nucleotide sequences of the type II *oriT*/*traM*/*traY* allele from ColB4-K98 and the type IV *traY* allele from R100-1 . .154
A. Introduction. 154
B. Materials and methods 156
  1. Bacterial strains and plasmids 156
  2. Plasmid purification 157
  3. Recombinant DNA techniques 157
  4. DNA sequencing 157
C. Results 157
  1. Cloning and sequencing strategy. 157
  2. *OriT* 159
  3. The *traM* protein 163
  4. The *traY* protein 164
D. Discussion. .165
E. Bibliography. 184

Chapter VIII. Nucleotide sequence of the surface

exclusion genes traS and traT from the IncFV plasmid pED208 188
A. Introduction. .188
B. Materials and methods 189
1. Bacterial strains and plasmids 189
2. Recombinant DNA techniques and DNA sequencing 189
3. Isolation of minicells 189
4. Surface exclusion assay and index. . . .189
C. Results 192
1. Cloning of the pED208 surface exclusion system 192
2. DNA sequencing studies 194
3. The 414 residue ORF. 195
4. TraS 197
5. TraT 198
6. The 244 amino acid ORF 200
7. Cross-reactivity between pED208 and F surface exclusion sysytems 201
D. Discussion. .202
E. Bibliography. 220

Chapter IX. Nucleotide sequence of the traYALE region of the IncFV plasmid pED208. 225
A. Introduction. .225
B. Materials and methods 228
1.Bacterial strains and plasmids. 228

2. Recombinant DNA techniques and sequencing 228
3. Isolation and labelling of minicells . .229

C. Results 230
1. Cloning and sequencing strategy. 230
2. The traY* gene of pED208. 232
3. The traA gene of pED208 233
4. The traL and traE genes of pED208 236
5. Polypeptides produced in minicells. 238

D. Discussion. 241

E. Bibliography. 260

Chapter X. Discussion. 266

A. Pilus biosynthesis. 266

B. Conjugal DNA metabolism 269

C. Surface exclusion 270

D. Regulation of Conjugation 271

E. Evolution of tra allele specificity 271

F. Bibliography. 275

Appendix. Publications 277

LIST OF FIGURES

Figure		Page
I.1	Genetic map of the transfer region of F plasmid.	15
II.1	Agarose gel electrophoresis of various restriction endonuclease digests of pED208.	40
II.2	Restriction endonuclease cleavage map of pED208.	42
II.3	_Hin_dIII/_Xho_I map of pED208 showing various clones and position of transposon insertions.	44
III.1	Restriction endonuclease cleavage map of ColB2.	69
III.2	Southern blots showing homology between ColB2 and F plasmids.	71
III.3	Nucleotide sequence of the ColB2 pilin gene.	73
IV.1	Competition ELISAs of purified F and ColB2 pili assayed with rabbit anti-pilus antisera.	97
IV.2	Competition ELISAs of amino terminal tryptic peptides treated with cyanogen bromide.	99
IV.3	Competition ELISAs of amino terminal synthetic peptides.	101
V.1	DNA sequencing strategy for _traM_, _finP_, _traJ_ _traY_ and _traYZ_ promoter from R1-19.	126
V.2	Nucleotide sequence of _oriT_, _traM_, _finP_, _traJ_ _traY_ and _traYZ_ promoter from R1-19.	128
V.3	Comparison of F and R1-19 _traM_ open reading frames.	130

V.4 Comparison of F and R1-19 *traY* open reading frames. 132

VI.1 Nucleotide sequence of five IncF *finP* genes and three *finP* mutations. 148

VI.2 Possible secondary structure of a type I *finP* RNA molecule. 150

VII.1 DNA sequencing strategy for *oriT*, *traM*, *finP*, *traY* and part of *traJ* from COlB4. 175

VII.2 Nucleotide sequence of *oriT*, *traM* and *traY* from ColB4 and *traY* from R100. 177

VII.3 Comparison of types I-IV *oriT* sequences. 179

VII.4 Comparison of types I-IV *traM* open reading frames. 181

VII.5 Comparison of F, R1, ColB4 and R100 *traY* open reading frames. 183

VIII.1 Restriction map of pED208 showing the location of the surface exclusion loci. 209

VIII.2 Restriction enzyme map of the *Sal*I/*Xho*I 3.7 kb fragment of pBF120 showing positions of subclones constructed. 211

VIII.3 Nucleotide sequence of the *Sal*I/*Xho*I 3.7 kb fragment of pBF120. 213

VIII.4 Predictive analysis of the secondary structure of the *traS* gene product. 215

VIII.5 Polypeptides produced in minicells containing pBF140, pBF141, and pBF120. 217

VIII.6 Comparison of TraTps from pED208, R100 and F. 219

IX.1 Map of pED208 showing the locations of IS2 and traYALE region. 251

IX.2 Sequencing strategy of pED208 traYALE. 253

IX.3 Nucleotide and translated sequence of traALE and part of traY. 255

IX.4 Comparison of pED208 and F traYALE open reading frames. 257

IX.5 Polypeptides produced in minicells containing pBF151 and pBF152. 259

LIST OF TABLES

Table		Page
II.1	Size and location of cloned PED208 fragments.	37
II.2	Phage sensitivity and transfer ability of various pED208 clones.	38
III.1	Complementation ability of chimeric plasmids containing F or ColB2 pilin gene.	65
III.2	Amino acid composition of tryptic peptides from ColB2 pilin.	66
III.3	Analysis of amino acids at the N-terminal and C-terminal of ColB2 pilin.	67
IV.1	Amino terminal tryptic peptide sequences of pED208, F and ColB2 pilins.	94
IV.2	Competition ELISA values obtained with synthetic and tryptic peptides of F and ColB2 pilin.	95
V.1	Galactikinase assay of F and R1-19 traYZ promoter activity.	124
VII.1	Percent homologies of oriT/traY/traM alleles from IncF plasmids.	173
VIII.1	Chimeras constructed from the pED208 plasmid.	243
VIII.2	Surface exclusion index of pED208 recombinants.	244

Chapter I

INTRODUCTION

Conjugation is the process whereby DNA is transferred from one bacterial cell to another. First described for the gram-negative species Escherichia coli (29), this process is widespread among bacteria, although the specific mechanisms involved vary considerably. Conjugal DNA transfer is a polar event, involving the passage of DNA from donor to recipient cells. The donor cell contains the genes necessary for this process, the transfer (tra) genes. These genes are encoded by large episomal elements termed conjugative plasmids.

Conjugative plasmids can be classified into incompatibility (Inc) groups, based on their ability to co-exist with other plasmids (45). Two plasmids which cannot both be maintained in one host cell belong to the same incompatibility group. Compatible plasmids (plasmids belonging to different Inc groups) often contain regions of DNA homology which correspond to related conjugation systems (16). An example of this is found in the IncFI-IV conjugative plasmids (46). Conjugation systems of IncF plasmids are the most extensively studied.

Common pili are usually more prevalent on the bacterial surfaces than conjugative pili, and are located either at the ends of a cell (polar pili) or distributed evenly over the cell's surface. Some types of somatic pili also confer a primitive form of movement to the cell known as twitching motility (11).

In the F system, all the genes necessary for conjugation are encoded by a 34 kilobase (kb) region on the F plasmid (56). Most of this region (33 kb) is contained within a single operon called the traYZ operon. Included in this operon are the genes necessary for pilus formation, mating pair stabilization, surface exclusion, and some of the genes involved in conjugal DNA transfer (Fig. I.1). Upstream of the traYZ operon is the site at which several control elements act to regulate transcription of this operon. Also upstream of this operon is the site of transfer initiation, oriT.

Several alleles have been defined for the genes involved in IncF plasmid transfer (55,56). The object of this work was to examine, at the molecular level, the similarities and differences of these genes and their predicted products. These comparisons can be used to determine the basis of allele specificity of tra genes. Furthermore, these sequences add to our knowledge of the structure and function of these gene products, increasing our understanding of the mechanisms involved in the transfer process. This chapter introduces the various processes involved in bacterial transfer of IncF plasmids.

A. Pilus Biosynthesis

The extent of piliation on cells containing IncF plasmids varies considerably. For example, cells harboring the F (IncFI) plasmid contain about 143 pili per 100 cells while pED208 (IncFV) produces about 1700 pili per 100 cells (21). Pili are composed of one repeating protein subunit, pilin (13,33), with an approximate molecular weight of 7000 (21,22). F pili are hollow tubular stuctures which are several micrometers in length, and have an outside diameter of 8 nm and an inner diameter of 2 nm. The pilin subunits have an alpha-helical content of 65 to 70% (9) and are arranged in a helical array (18) within the pilus.

F-like pili are required for adsorption of male-specific bacteriophages. F-specific single-stranded DNA phages are filamentous structures which adhere to the tip of a pilus whereas pilus specific RNA phages attach to the side of this structure (14,15). Variations in levels of phage sensitivity occur within cells harboring different IncF plasmids (8,42). For example, cells containing pED208 are resistant to F-specific RNA phages such as R17 (12) but they are sensitive to DNA phages such as f1.

Another method of differentiating F-like pili is based upon their serological specificity (21,28). This method identified five serolgical groups for IncF pili. F and pED208 pilins both contain N-acetylated amino termini

(20,22), and, in the case of pED208, this acetylated amino terminus and first five residues constitute the major antigenic determinant of the protein (58). Based on these criteria, Willetts and Maule (55) identified four F-like pilus alleles, and Frost *et al*. added a fifth (pED208) (21).

Despite the fact that pili appear relatively simple, being composed of the single repeating subunit pilin, at least fourteen genes are required for F pilus assembly (56). These include *traA*,*L*,*E*,*K*,*B*,*P*,*V*,*W*,*C*,*U*,*F*,*Q*,*H* and an amino terminal domain of *traG*. These proteins are encoded within the first 20 kb of the *tra* operon (Fig. I.1). *TraA* is the structural gene for F pilin (33), encoding a processed protein of 70 amino acids and a 51 residue leader sequence (22). The product of F *traQ* is required for conversion of F *traA* product to pilin (25,35), possibly by a proteolytic modification. It has been shown that *traG* product is required for immunologically active F pilin production (27), suggesting that the *traG* product may be involved in the amino-terminal acetylation. Although *traL* and *traE* have been sequenced (22), their functions remain undefined, as do the remainder of the pilus assembly genes. Most of these pilus assembly proteins are associated with the cell envelope (55) and these products are presumably involved in the modification and transport of pilin, or in forming a basal structure upon which pili can polymerize (or retract). Polymerization is believed to be an energy-dependent process (38).

Another characteristic of F-like pili is their ability to retract (37). Once initial cell-cell contacts are established between the donor pilus and recipient cell, this retraction process is thought to bring the two cells into close proximety, forming unstable mating pairs (reviewed by Silverman [47]). The site of interaction on the recipient cell wall with the pilus tip remains undefined, although some information suggests that ompA protein in the recipient cell may be involved (55).

B. Mating Pair Stabilization

Once the donor pilus has retracted and unstable mating pairs have formed, the next event necessary for conjugation is the formation of stable (shear-resistant) mating pairs. Mutations in traN and the latter part of traG produce unstable mating aggregates, but never progress to the stable cell-contact stage (32).

The exact mechanism of stable mating-pair formation remains unresolved, although recipient bacteria lacking the outer membrane protein ompA are unable to form these stable interactions (48). The requirement of ompA suggests that membrane interactions may occur, perhaps forming a pore or channel through which DNA transfer can occur (57).

Panicker and Minkley (41) demonstrated that transfer does not occur through an extended F pilus but requires surface contact between donor and recipient cells. Once these interactions have occured, addition of low levels of the detergent sodium dodecyl sulphate (SDS) have no affect upon conjugal transfer (5) even though this treatment removes pili (52). Addition of SDS before mating-pair stabilization has occurred blocks transfer, indicating that pili are required for the initial cell-cell interaction, but not for the actual DNA transfer process.

C. Conjugal DNA Metabolism

Upon completion of stable mating pair formation, a single pre-existing strand of double-stranded donor plasmid DNA is passed from the donor cell to the recipient cell with the 5' terminus leading. For IncF plasmids, this passage requires several plasmid-encoded genes which include traD, traI, traM, traY, traZ and oriT whose roles are discussed below (reviewed by Willetts and Wilkins [57]). Concurrent with transfer is the synthesis of a replacement strand in the donor cell and of a complimentary strand in the recipient cell occurs. These processes require host encoded de novo primer synthesis and DNA polymerase III holoenzyme.

OriT is the site at which conjugal DNA transfe initiates. In F, this site is located at one end of the tra region, upstream of the traYZ operon (Fig. I.1, [53]). The polarity of transfer is such that the last region of DNA to enter the recipient cell is the tra operon. The nucleotide sequence of oriT from two IncF plasmids (F, R1) has been established (40,51). Thompson et al. also identified three oriT-specific nick sites in F by using a lambda-oriT system (17) which are conserved in R1 (IncFII). These nicks contained a hydroxyl group at the 3' terminus but could not be religated with T4 ligase, indicating that the 5' terminus may be modified or a small number of bases removed. Also contained within oriT are several sequences that are recognized by tra proteins such as the traY and perhaps traM products. Although these recognition sequences have not been identified, they serve to define at least four oriT alleles in the IncF plasmids, based on complementation analysis (55). OriT sequences contain several direct and inverted repeats and a large adenine-thymine rich region. It is believed that these sequences play a role in tra protein recognition of oriT (40).

The enzyme that is probably responsible for the reversible nicking and ligating activity at oriT is the traYZ endonuclease (17). The two genes required for this activity (traY and traZ) are located at opposite ends of the tra operon (Fig.I. 1). Willetts and Maule (55) have demonstrated that traZ products but not traY products from R1 and F are interchangable. Since the oriT nicking sites of these two plasmids are conserved (see above), these authors proposed that the traZ component of the traYZ endonuclease provides nicking activity while the traY product is reponsible for plasmid-specific sequence recognition at oriT.

This enzyme is thought to be located at the interface of the cytoplasm and cell membrane, based on the cellular locations of the traY (cell envelope) and traZ (cytoplasm) products (reviewed by Willetts and Wilkins [55]). This location would "anchor" oriT near the cell membrane, presumably near a pore through which the transfered strand will pass. The observed oriT nicking and religation activity appears to be in equilibrium within the donor cell (17). This activity is seen regardless of the "conjugal" state of the donor, provided traY and traZ products are present. Upon initation of conjugal transfer, the traYZ endonuclease may be dislodged from oriT and replaced by other tra proteins (57).

DNA transfer is initiated once stable mating pairs have formed. The signal which triggers this response is undefined, but a likely candidate is the traM protein. This protein is not required for pilus assembly, stable mating pair formation or oriT nicking but is essential for DNA transfer (26). The F traM product is located in the inner membrane (6), and has been sequenced (50). Willetts and Maule (55) identified four traM alleles amongst the twelve IncF plasmids tested, corresponding to the four oriT alleles.

Once the signal to begin conjugal transfer has been received, the two strands of donor plasmid DNA must be unwound to allow transmission of a single strand into a recipient cell. The product of traI is a DNA helicase I (2) and presumably provides this activity. This protein is located in the cytoplasm (1) and posseses a DNA-dependant ATPase activity which supplies the energy needed for DNA unwinding. If the traI product is immobilized near the proposed pore through which strand transfer may take place, one can imagine that the traI activity would separate the two plasmid strands, propelling the transfered strand into the recipient cell (47). Willetts and Maule (55) have determined that traI products of IncF plasmids are not interchangable.

By using a temperature-sensitive F traD mutation, Panicker and Minkley demonstrated that the traD product, which is required for DNA transfer, acts at a stage after mating pair formation (41).

D. Surface Exclusion

Transfer of a plasmid into a cell which already harbors a copy of that plasmid is an unproductive event. Conjugative plasmids encode means to prevent this redundant mating, a phenomenon referred to as surface exclusion (30). IncF plasmids contain two surface exclusion genes (traS and traT) whose products provide exclusion by two separate and independant mechanisms. These genes are located within the traYZ operon, immediately downstream of the region required for pilus biosynthesis and are followed by traG and traD (Fig. I.1). Willetts and Maule (55) identified four surface exclusion alleles in F-like plasmids, but did not determine if this specificity was dictated by traS, traT or both genes.

The product of traS is an inner membrane protein with a molecular weight of 18,000 daltons in F (4), although this protein varies in size in different IncF plasmids. TraS protein is capable of inhibiting conjugal DNA transfer even when stable mating pairs have formed suggesting that the traS product may inhibit triggering of conjugal DNA metabolism (31). The exact mechanism of traS action is not understood.

The product of F traT is a 25,000 dalton outer membrane protein (4). This protein is produced in large quantities, with at least 20,000 copies present per cell. Achtman et al. (5) demonstrated that this protein provides surface exclusion by preventing the formation of stable mating aggregates. It has been suggested that the pilus tip may interact with traT protein on the "recipient" cell's surface, blocking an interaction essential for stable mating-pair formation (54). In support of this hypothesis, Willetts and Maule found that in most cases pilus type alleles matched exclusion alleles. The discrepencies may be attributable to traS allele specificity. The primary structure of R100 traT has been published (38). Perumal and Minkley (43) found that F traT product is a lipoprotein. TraT products have also been implicated in mediating serum resistance (34) and enhancing resistance of *E. coli* to phagocytosis (7).

E. Regulation of Transfer

Since most of the proteins required for transfer are encoded within the 33 kd traYZ operon, regulation of the traYZ promoter (P_{YZ}) would control transfer. This is precisely the control mechanism used in F-like plasmids. Three genes are involved in P_{YZ} control: traJ, finO and finP. The product of traJ is required for transcriptional activation of the P_{YZ} promoter, and in the F system causes a four fold increase in this promoter's transcriptional levels (19,36). The nucleotide sequence of F traJ has been determined (19), and Willetts and Maule (55) identified three traJ/P_{YZ} alleles in the IncF plasmids. Mullineaux and Willetts (36) have localized the

P_{YZ} promoter, and found that this promoter does not share much homology with the consensus promoter sequence in procaryotes (23). This is not surprising as positively controlled promoters are usually unique sequences (44).

In turn, *traJ* is regulated by a *f*ertility *in*hibiton system encoded by *finO* and *finP*. Products of these two genes act in concert to lower *traJ* transcription levels about three fold (36). *FinP* is a small gene overlapping the *traJ* promoter, being transcribed in the opposite direction to *traJ* (Fig. I.1, [36]). Although F *finP* encodes a putative 24 residue peptide, Mullineaux and Willetts suggested that the *finP* product is an RNA molecule that, together with *finO* product, interacts with its complementary sequence to prevent *traJ* transcription and/or translation. Changes in the RNA sequence would account for *finP* plasmid specificity, as six IncF *finP* alleles have been identified (55).

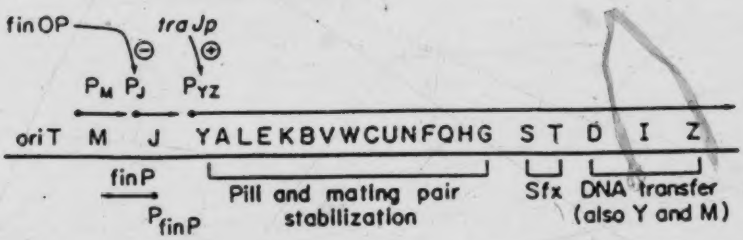
IncF *finO* products are interchangable, with two classes of *finO* being reported (55). The lack of either *finO* or *finP* product results in increased *traJ* transcription and thus increased levels of *traYZ* operon transcription. A plasmid in this state is referred to as derepressed. F is a naturally occuring derepressed plasmid as it does not encode a functional *finO* gene product. However, if a compatable FinO$^+$ plasmid is present in the same cell as F, the F *tra* operon is repressed.

Derepression results in approximately a hundred fold increase in the number of cells capable of conjugation.

The F genes traM, traI and finP have their own promoters which provide constitutive expression of these genes and are thus independant of traJ/finOP regulation (55). The R100 traT gene appears to encode its own promoter (39), but this is not the case for F traT (36). Plasmids from other incompatibiltiy groups have different fertility inhibition systems which are reviewed by Willetts and Skurray (56).

FIGURE I.1

Genetic map of the transfer region of F plasmid.


finOP
traJp
⊖
⊕
P_M
P_J
P_{YZ}
oriT
M
J
YALEKBVWCUNFQHG
S
T
D
I
Z
finP
P_{finP}
Pili and mating pair stabilization
Sfx
DNA transfer (also Y and M)

Genetic map of F conjugation region. Distance from ori T to *tra Z* is −33 kb.

E. BIBLIOGRAPHY

1. **Abdel-Monem, M., H. -F. Lauppe, J. Kartenbeck, H. Durwald, and H. Hoffmen-Berling**. 1977. Enzymatic unwinding of DNA. III. Mode of action of Escherichia coli DNA unwinding enzyme. J. Mol. Biol. 110:667-685.

2. **Abdel-Monem, M., G. Taucher-Scholz, and M.-Q. Klinkert**. 1983. Identification of Escherichia coli DNA helicase I as the traI gene product of the F sex factor. Proc. Natl. Acad. Sci. U.S.A. 80:4659-4663.

3. **Achtman, M., N. Kennedy, and R. Skurray**. 1977. Cell-cell interactions in conjugating E. coli: role of traT protein in surface exclusion. Proc. Natl. Acad. Sci. U.S.A. 74:5104-5108.

4. **Achtman, M., and R. Skurray**. 1977. A redefinition of the mating phenomenon in bacteria. In Microbial Interactions, Receptors and Recognition. Ser. B. 3:232-279. ed. J.L. Reissig, London: Chapman & Hall.

5. **Achtman, M.G., G. Morelli, and S. Schwachow**. 1978. Cell-cell interactions in conjugating Eschericia coli: role of F pili and fate of mating aggregates. J. Bacteriol. 135:1053-1061.

6. **Achtman, M., P.A. Manning, C. Edelbluth, and P. Heelich**. 1979. Export without proteolytic processing of inner and outer membrane proteins encoded by F sex factor tra cistrons in Escherichia coli minicells. Proc. Natl. Acad. Sci. U.S.A. 76:4837-4841.

7. **Aguero, M.E., L. Aron, A.G. Deluca, K.N. Timmins, and F.C. Cabello**. 1984. A plasmid-encoded outer membrane protein, *traT*, enhances resistance of *Escherichia coli* to phagocytosis. Infec. Immun. **46**:740-746.

8. **Alfaro, G., and N. Willetts**. 1972. The relationship between the transfer systems of some bacterial plasmids. Genet. Res. **20**:279-289.

9. **Armstrong, G.D., L.S. Frost, P.A. Sastry, and W. Paranchych**. 1980. Comparative biochemical studies on F and EDP208 conjugative pili. J. Bact. **141**:333-341.

10. **Beachey, E.H.**, ed., 1980. Bacterial Adherance. Chapman and Hall, London and New York, 464 p.

11. **Bradley, D.E.** 1980. A function of pili on *Pseudomonas aeruginosa* PAO polar pili: Twitching motility. Can. J. Microbiol. **26**:146-154.

12. **Bradley, D.E., J.N. Coetzee, T. Bothma, and R.W. Hedges**. 1981. Phage F_o*lac*: an F_o*lac* plasmid-dependent bacteriophage. J. Gen. Microbiol. **126**:405-411.

13. **Brinton, C.C.** 1971. The properties of sex pili, the viral nature of conjugal genetic transfer systems, and some possible approaches to the control of drug resistance. Critical Rev. Microbiol. 1:105-160.

14. **Caro, L.G. and M. Schnoss**. 1966. The attachment of male-specific bacteriophage f1 to sensitive strains of *E. coli*. Proc. Natl. Acad. Sci. **56**: 126-132.

15. **Crawford, E.M. and R.F. Gesteland**. 1964. The adsorption of bacteriophage R17. Virology. **22**:165-167.

16. **Datta, N.** 1975. Epidemiology and classification of plasmids. In Microbiology 1974, ed. D. Schlessinger, pp 9-15. Am. Soc. Micro., Washington, D.C.

17. **Everett, R., and N. Willetts.** 1980. Characterization of an in vivo system for nicking at the origin of conjugal DNA transfer of the sex factor F. J. Mol. Biol. **136**:129-150.

18. **Folkhard, W., K.R. Leonard, S. Malmsey, D. A. Marvin, J. Dubochet, A. Engel, M. Achtman, and R. Helmuth.** 1979. X-ray diffraction and electron microscope studies on the structure of bacterial F pili. J. Mol. Biol. **130**:145-160.

19. **Fowler, T., L. Taylor, and R. Thompson.** 1983. The control region of the F plasmid transfer operon: DNA sequence of the traJ and traY genes and characterization of the traYZ promoter. Gene **26**:79-89.

20. **Frost, L.S., G.D. Armstrong, B.B. Finlay, B.F.P. Edwards, and W. Paranchych.** 1983. N-terminal amino acid sequencing of EDP208 conjugative pili. J. Bacteriol. **153**:950-954.

21. **Frost, L.S., B.B. Finlay, A. Opgenorth, W. Paranchych and J. Lee.** 1985. Characterization and sequence analysis of pilin from IncF plasmids. J. Bacteriol. **164**:1238-1247.

22. **Frost, L.S., W. Paranchych, and N.S. Willetts.** 1984. DNA sequence of the F traALE region that incudes the gene for F pilin. J. Bacteriol. **160**:395-401.

23. **Hawley, D., and W. McClure.** 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res. 11:2237-2255.

24. **Helmuth, R., and M. Achtman.** 1978. Cell-cell interactions in conjugating Escherichia coli: Purifictation of F pili with biological activity. Proc. Natl. Acad. Sci. U.S.A. 75:1237-1241.

25. **Ippen-Ihler, K., D. Moore, S. Laine, D.A. Johnson and N.S. Willetts.** 1984. Synthesis of the F-pilin polypeptide in the absence of F traJ product. Plasmid. 11:116-129.

26. **Kingsman, A., and N. Willetts.** 1978. The requirements for conjugal DNA synthesis in the donor strain during Flac transfer. J. Mol. Biol. 122:287-300.

27. **Laine, S., D. Moore, P. Kathir, and K. Ippen-Ihler.** 1985. Genes and gene products involved in the synthesis of F-pili, p. 535-553. In D. Helinski, S. Cohen, D. Clewell, D. Jackson, and A. Hollaender (eds.), Plasmids in Bacteria. Plenum Publishing Company, New York.

28. **Lawn, A.M. and E.E. Meynell.** 1970. Serotypes of sex pili. J. Hygiene 68:683-694.

29. **Lederberg, J., and E. Tatum.** 1946. Gene recombination in Escherichia coli. Nature 158:558.

30. **Lederberg, J., L.L. Cavalli, and E.M. Lederberg.** 1952. Sex compatibility in Escherichia coli. Genetics 37:720-730.

31. **Manning, P.A., and M. Achtman.** 1979. Cell to cell

interactions in conjugating Escherichia coli: the involvement of the cell envelope. In M Inouye (ed.), Bacterial outer membrane biogenesis and functions, pp.409-447. New York. Wiley.

32. **Manning, P.A., G. Morelli, and M. Achtman.** 1981. traG protein of the F sex factor of Escherichia coli K-12 and its role in conjugation. Proc. Natl. Acad. Sci. U.S.A. **78**:7487-7491.

33. **Minkley, E.G., S. Polen, C.C. Brinton and K. Ippen-Ihler.** 1976. Identification of the structural gene for F pilin. J. Mol. Biol. **108**:111-121.

34. **Moll, A., P.A. Manning, and K.N. Timmis.** 1980. Plasmid-determined resistance to serum bactericidal activity: a major outer membrane protein, the traT gene product, is responsible for plasmid-specified serum resistance in Escherichia coli. Infect. Immun. **28**:359-367.

35. **Moore, D., B.A. Sowa, and K. Ippen-Ihler.** 1982. A new activity in the Ftra operon which is required for F-pilin synthesis. Mol. Gen. Genet. **188**:459-464.

36. **Mullineaux, P., and N. Willetts.** 1984. Promoters in the transfer region of plasmid F. In D. Helinski, S. Cohen, D. Clewell, and A. Hollaender (eds.), Plasmids in Bacteria, pp.605-614, Plenum Press, New York.

37. **Novotny, C.P., and P. Fives-Taylor.** 1974. Retraction of F pili. J. Bacteriol. **117**:1306-1311.

38. **O'Callagan, R.J., L. Bundy, R. Bradley, and W. Paranchych**. 1973. Unusual arsenate poisoning of the F pili of *Escherichia coli*. J. Bacteriol. **115**:76-81.

39. **Ogata, R., C. Winters, and R.P. Levine**. 1982. Nucleotide sequence analysis of the complement resistance gene from plasmid R100. J. Bacteriol. **151**:819-827.

40. **Ostermann, E., F. Kricek, and G. Hogenauer**. 1984. Cloning the origin of tranfer region of the resistance plasmid R1. EMBO J. **3**:1731-1735.

41. **Panicker, M.M., and E.G. Minkley, Jr**. 1985. DNA transfer occurs during a cell surface contact stage of F sex factor-mediated bacterial conjugation. J. Bacteriol. **162**:584-590.

42. **Paranchych, W**. 1975. Attachment, ejection and penetration stages of the RNA phage infectious process. *In* RNA Phages. N. Zinder (ed.), pp.85-111. Cold Spring Harbor, NY: Cold Spring Harbor Lab.

43. **Perumal, N., and E.G. Minkley**. 1984. The product of the F sex factor *traT* surface exclusion gene is a lipoprotein. J. Biol. Chem. **259**:5357-5360.

44. **Raibaud, O. and M. Schwartz**. 1984. Positive control of transcription initiation in bacteria. Ann. Rev. Genet. **18**:173-206.

45. **Scott, J.R**. 1984. Regulation of plasmid replication. Microbiol. Rev. **48**:1-23.

46. **Sharp, P.A., S.N. Cohen, and N. Davidson**. 1973.

Electron microscope heteroduplex studies of sequence relations among plasmids of *E. coli* II. Studies of drug resistant (R) factors and F factors. J. Mol. Biol. **75**:235-255.

47. **Silverman, P.M.** 1986. The structural basis of prokaryotic DNA transfer. *In* Bacterial outer membranes as model systems. In press.

48. **Skurray, R. A., R.E.W. Hancock, and P. Reeves.** 1974. Con$^-$ mutants: class of mutants in *Escherichia coli* K-12 lacking a major cell wall protein and defective in conjugation and adsorption of a bacteriophage. J. Bacteriol. **119**:726-735.

49. **Swaney, L.P., Y. -P. Liu., K. Ippen-Ihler, and C.C. Brinton, Jr.** 1977. Genetic complementation analysis of *Escherichia coli* type I somtatic pilus mutants. J. Bacteriol. **130**:506-511.

50. **Thompson, R., and L. Taylor.** 1982. Promoter mapping and DNA sequencing of the F plasmid transfer genes *traM* and *traJ*. Mol. Gen. Genet. **188**:513-518.

51. **Thompson, R., L. Taylor, K. Kelly, R. Everett, and N. Willetts.** 1984. The F plasmid origin of transfer: DNA sequence of wild-type and mutant origins and location of origin-specific nicks. EMBO J. **3**:1175-1180.

52. **Tomoeda, M., M. Inuzuka, and T. Date.** 1975. Bacterial sex pili. Prog. Biophys. Mol. Biol. **30**:23-56.

53. **Willetts, N.S.** 1972. Location of the origin of transfer of the sex factor F. J. Bacteriol. 112:773-778.

54. **Willetts, N., and J. Maule.** 1974. Interactions between the surface exclusion systems of some F-like plasmids. Genet. Res. **24**:81-89.

55. **Willetts, N.S., and J. Maule.** 1986. Specificities of IncF plasmid conjugation genes. Genet. Res. **47**:1-11.

56. **Willetts, N.S., and R. Skurray.** 1980. The conjugation system of F-like plasmids. Ann. Rev. Genet. **14**:41-76.

57. **Willetts, N., and B. Wilkins.** 1984. Processing of plasmid DNA during bacterial conjugation. Microbiol. Rev. **48**:24-41.

58. **Worobec, E. A., W. Paranchych, J. M. R. Parker, A. K. Taneja, and R. S. Hodges.** 1985. Antigen-antibody interaction: the immunodominant region of EDP208 pili. J. Biol. Chem. **260**:938-943.

Chapter II

CHARACTERIZATION OF CONJUGATIVE PLASMID pED208[1]

A. INTRODUCTION

Many conjugative plasmids are known and have been assigned to incompatibility groups. Characteristic of conjugative plasmids is the ability to encode all proteins required for the transfer of the plasmid. F is the classic conjugative plasmid which has been most extensively studied (for a review see reference 19). pED208 (also called EDP208) is a derepressed derivative of a naturally occuring *lac* plasmid, Fo*lac*, which was originally identified in *Salmonella typhi* (9). pED208 pili are serologically distinct from F-like conjugative pili, and they are resistant to F-specific RNA phages (4). However, pED208 pili do confer sensitivity to the F-specific filamentous DNA phages (3).

This chapter describes the genetic characterization of plasmid pED208. Restriction endonuclease mapping and transposon mutagenesis were used to determine the size of

[1] A version of this chapter has been published. **Finlay, B.B., W. Paranchych, and S. Falkow.** 1983. J. Bacteriol. 156:230-235.

the plasmid and the location of the region encoding pilus formation. In addition, the cloning of four large *Hin*dIII fragments allowed the localization of the regions encoding *lac*, *oriV*, *oriT*, and *tra*.

B. Materials and Methods

1. Bacterial Strains.

Plasmid pED208 was carried in the host strain JC6256 (*Escherichia coli* K-12 F$^-$ *trp lac*) and was kindly donated by N.S. Willetts, Department of Molecular Biololgy, University of Edinburgh. Clones were transformed into HB101 (F$^-$ *thr leu thi pro hsr hsm recA lac gal* Str^r), NM259 (*met gal lacY supE supF hsdM hsdR*), or AA125 (*his trp bio* Str^r). SF800 (*thy polA* Nal^r) (8,10) was used to determine *oriV*. pMR5 was used as the Tn1 donor (15). Tn5 was chromosomally contained in DB89-007 (F$^-$ *trpE5 ilvA-007*::Tn5) and was kindly donated by D.E. Berg, Department of Microbiology, Washington University School of Medicine. ED24 (F$^-$ *lac* Str^r Spc^r) was used as a Spc^r recipient for pED208::Tn1 and pED208::Tn5.

2. Isolation of Plasmid DNA.

The cleared lysate method described by Crosa and Falkow (6), followed by CsCl density centrifugation was used to obtain most DNA. pED208::Tn5 DNA was isolated by the method developed by Birnboim and Doly (2). pED208::Tn1 DNA was isolated as described by Portnoy et al. (14).

3. Restriction Endonuclease Digests.

Purified plasmid DNA (0.5 micrograms) was digested with *BamHI*, *HindIII*, *SstI*, *XhoI*, or *PvuI* (Bethesda Research Laboratories, Gaithersburg, Md.) in buffers, as specified by Davis et al. (7), at 37°C for 1 h. When two enzymes required different buffers, digestion in one buffering system was performed, and the DNA was ethanol precipitated and suspended in the second buffering system and digested. Cleaved DNA was subjected to electrophoresis in a horizontal 0.6% agarose gel, using 0.04 M Tris, 0.02 M sodium acetate, 0.002 M EDTA (pH 8) as a buffer. Lambda DNA that was digested with *HindIII* was electrophoresed also and used as molecular weight markers.

4. Fragment isolation.

DNA (1 microgram) was cleaved with an appropriate restriction endonuclease, applied to a 0.7% agarose gel, and electrophoresed in a Tris-borate buffer (13). After ethidium bromide staining (0.5 mg/liter), the gel was illuminated on a long wavelength UV light box, and the desired band was excised. This was placed in dialysis tubing and electrophoresed in the 1/20 Tris-acetate buffer mentioned previously. The DNA in the supernatant solution was ethanol precipitated and suspended in the desired buffer.

5. Recombinant DNA Techniques.

a. Cloning _Hin_dIII fragments (pBF101-105). A 0.2-microgram amount of the appropriate restriction fragment was isolated from a gel and combined with a fivefold excess of vector DNA (pBR322 [18]) or pACYC184 [5]) which had previously been digested with _Hin_dIII, treated with calf alkaline phosphatase, heated to 65°C for 10 min, phenol extracted, and ethanol precipitated. Fragment and vector DNAs were suspended in 20 microliters of ligase buffer (as specified by Boehringer Mannheim Corp., New York, N.Y.) and one unit of ligase (Boehringer Mannheim) was added. Ligation was carried out for two to three hours at 15°C. Competent cells were made by suspending a mid-log-phase culture in 0.1 M $CaCl_2$. These were added to the ligation mixture at 0°C for 45 min, heat shocked (42°C, 10 min), diluted with an equal volume of L-broth (37°C, 2h, with shaking) and plated onto ampicillan (if pBR322) or chloramphenicol (if pACYC184) plates. Insertions were detected by sensitivity to tetracycline.

b. Cloning _Hin_dIII-_Xho_I fragment (pBF106). pBR322 was digested with _Hin_dIII and _Sal_I, and the large vector fragment was isolated. This was combined with the

HindIII-XhoI 12-kilobase (kb) fragment. Since XhoI and SalI have complementary ends, the ligation of this mixture was performed as described above, and Amp$^+$ Tc$^-$ colonies were selected, (The XhoI-SalI site is not regenerated.)

c. **Cloning XhoI fragment (pBF111).** pBF101 was digested with XhoI, which liberates only two XhoI fragments. These were ligated into pBR322/SalI, and Amp$^+$ Tc$^-$ colonies were selected. The 7.5-kb insertions represented pBF111.

6. f1 Phage Sensitivity Tests.

Lawns of strains to be tested were seeded on L-agar plates. A one-microliter amount of a 7.4×10^{10} particle/ml stock solution of f1 phage was placed on the surface, and the plate was incubated at 37°C. Cultures were considered f1 sensitive if a cleared spot appeared and f1 resistant if no clearing was seen.

7. pED208::Tn1 Mutagenesis.

pMR5, a Tn1-containing plasmid (15), and pED208 were mated into JC6256. pMR5 carries kanomycin resistance, Tn1 encodes ampicillan resistance, and pED208 is lac$^+$. Kan$^+$ Amp$^+$ lac$^+$ colonies were selected. These were f1 phage sensitive. Cultures of 1 ml were grown at 30°C for 12 h and then diluted 1:50. This was repeated for 3 days. The final culture was mixed with an equal volume of ED24 (chromosomal Spc^r), and 2 ml of L-broth was added. Mating proceeded for 3 h at 30°C, and 0.1 ml of the mating mixture was then plated onto L1 Spc Amp plates. It was reasoned that although Tn1 may have inserted into regions required for transfer, recent mutants would still have intact transfer proteins and should be able to mate once. Five f1 phage-resisitant colonies were isolated from 200 colonies. These were grown at 42°C with shaking, for 16 h, to cure pMR5, as pMR5 is unstable at this temperature (15). All colonies derived from the cultures were Kan^s, indicating that they were cured of pMR5. Plasmid DNA was prepared from these mutants and characterized.

8. pED208::Tn5 Mutagenesis.

pED208 was mated into DB89-007 or DB89-301 (chromosomal Tn5 [Kan^r]) and grown at 37°C to late-log phase. This culture was mated with ED24 and plated onto L1 Spc Kan plates. Only ED24 (Spc^r) organisms that contain pED208::Tn5 (Lac$^+$ Kan^r) grew. These were screened for f1 phage sensitivity, and 37 out of 250 colonies were f1 resistant. Plasmid DNA from these organisms was isolated and characterized.

9. Mating Tests.

Plasmid clones to be tested for transfer fuction were transformed into AA125. The derivative strains were grown to mid-log phase and mixed with an equal volume of the recipient ED24 (Spc^r). An equal volume of fresh L-broth was added, and the cells were incubated at 37°C for 4 h. The mixture was then plated onto Spc plates which contained either ampicillin or chloramphenicol, depending on the vectors present. Only cells which were able to transfer the clone to ED24 were able to grow. The growth of Spc^r Amp^r (or Spc^r Cm^r) transconjugates was then an assay of the tranfer proficiency of the plasmid tested.

C. RESULTS AND DISCUSSION

1. Restriction Enzyme Analysis of pED208.

Isolated pED208 plasmid DNA was digested wtih five restriction enzymes-BamHI, HindIII, XhoI, SstI, and PvuI. These digests produced 6, 7, 3, 7, and 17 fragments, respectively (fragments smaller than 1 kb were not mapped). Digestions were performed with either one or two of these enzymes present in all possible combinations. The restriction fragments were separated on a 0.6% Tris-acetate agarose gel and visualized after ethidium bromide staining (Fig. II.1). The fragments were sized by their relative mobilities compared with lambda/HindIII DNA. A summation of the sizes of each set or restriction fragments suggested that plasmid pED208 is about 90 kb in length (60 x 10^6 daltons). This is the first reported size of plasmid pED208.

Digestion with single endonucleases and mixtures of endonucleases was performed similarly to the method described by Smith et al. (17) to obtain a restriction map. To complete the map, we isolated large fragments from vertical Tris-borate gels (see above) and subjected them to secondary digestions. After resolution by gel electrophoresis, enough information was obtained to generate a map of pED208 (Fig. II.2). The _Pvu_I site in brackets at 45 kb is flanked by two fragments of 8.7 and 3.5 kb. However, the orientation of these fragments was not determined.

From restriction digests, pED208 is approximately the same size as other _E. coli_ conjugative plasmids, such as F (100 kb [19]) and ColB2 (90 kb; Chapter III). We also found an asymmetric restriction site distribution in pED208, with many more sites in the 75- to 7-kb region then in the remainder of the plasmid, which inludes the region encoding pilus production. This was also observed with the _Eco_RI endonuclease (data not shown). It has been suggested that transfer regions of conjugative plasmids contain fewer restriction endonuclease sites in regions required for transfer (N. Willetts, personal communication; B. Finlay, unpublished data). This could possibly provide survival advantages for interstrain mating of conjugative plasmids since retriction sites in essential areas of the plasmid (such as tra) would likely be undesirable.

2. Transposon Mutagenesis to Identify the Pilus-Producing Region.

Mutagenesis with transposons Tn1 and Tn5 was used to identify the region of DNA necessary for pilus formation. Pilus formation in mutagenized strains was tested by screening for resistance to the filamentous DNA phage f1.

If transposon insertions occurred in a region essential for pilus production, phage-resistant pED208::Tn strains would be obtained. Six pED208::Tn1 phage-resistant mutants were isolated from 200 pED208::Tn1 colonies. (Lac$^+$ Amp^r Spc^r) (3%). Once they were cured of the Tn1 donor plasmid (pMR5) at 42°C, only one plasmid which had a molecular weight of about 95 Kb (pED208 plus Tn1) was present. Tn1 has one BamHI site and one PvuI site and is very similar to Tn3 (11). (However, it was found that although Tn3 has an SstI site Tn1 does not.) By using BamHI and PvuI digestions of pED208::Tn1, we determined the site of insertion and the orientation of the transposon, since new restriction sites would be introduced upon transposon insertion (Fig. II.3). Of the six Tn1 mutants, four were simple insertions, one inserted twice, and one had a 15.8 Kb deletion at the IR-L (left inverted repeat) of Tn1. All mutants were within a 16.5-Kb region.

Since double insertions and deletions were more difficult to map, pED208::Tn5 phage resistant mutants were also made. Of 250 pED208::Tn5 colonies (Lac$^+$ Kan^r Spc^r), 37 (14.8%) were f1 phage resistant. Tn5 has one central BamHI site, two HindIII sites and two XhoI sites near its termini (16). By digestion with BamHI, HindIII, or XhoI, we determined the sites of Tn5 insertions (Fig. II.3). All 37 mutants were single insertions. These mutants all fell within a 16-kb region, the same region as the Tn1 mutants. The significance of the clustering of Tn5 insertions around the HindIII site at 34 kb is explored in Chapter IX.

From these transposon mutant studies, it is apparent that only one 16.5-kb region of the plasmid (19.5 to 36 kb) was required for pilus formation. This is comparable in size to the region of the F pilus operon required for F pili formation (19.5 kb [19]).

3. Location of Lac, OriV, and Pilus-Producing Regions of the Plasmid.

Four large HindIII fragments (27.5, 23, 20.5, 12 kb) were each ligated into a HindIII-cleaved calf-alkaline phophatase-treated vector, pACYC184. These large fragments were preparatively isolated from an pED208/HindIII digest separated by 0.7% agarose gel electrophoresis. The fragments were then ligated into pACYC184 and transformed in HB101. Plasmids from Cm^r Tc^s strains were isolated, and a representative clone of each fragment was selected (Table II.1).

Collectively, these four clones contain 93% of the plasmid DNA. To determine which DNA fragment contained the *lac* region, bacteria containing the four chimeras were plated onto MacConkey agar and scored for Lac$^+$ phenotype. Only pBF103 was Lac$^+$, whereas pBF101, pBF102, and pBF104 were Lac$^-$. This places the *lac* region of pED208 in the 20.5-kb fragment (position 57 to 77.5).

To locate the origin of vegetative replication (*oriV*), pBF101, pBF102, pBF103, and pBF104 were each transformed into SF800, a *polA*$^-$ strain, since small multicopy vectors, such as pACYC184 and pBR322, have an *oriV* which requires *polA* to replicate and are unable to grow in SF800. However, large conjugative plasmids such as pED208 have a *polA*-independant *oriV* and are capable of growth in SF800. If a clone contained such an *oriV* from pED208, it would replicate and grow in SF800. Only SF800 (pED208) and SF800 (pBF103) were able to grow, whereas SF800 (pBF101), SF800 (pBF102), SF800 (pBF104) and SF800 (pACYC184) were unable to do so. This suggests that the region required for vegetative replication of pED208 is contained within the 20.5-kb *Hin*dIII fragment, from 57 to 77.5 kb. As shown above, this fragment also contains the *lac* region.

Since pBF101-pBF104 represents about 93% of the DNA of pED208, it was hoped that one of these fragments would contain the genes required for pilus production. AA125 (pBF101) was found to produce pili as judged by f1 phage sensitivity. Electron microscope examination of these cells confirmed that they produce normal amounts of pili. All other colonies were phage resistant, including two subclones of pBF101, pBF106 and pBF111 (Table II.2). This suggests that all of the genes required for pilus synthesis are present in pBF101. This DNA fragment includes the region shown to be necessary for pilus production by transposon mutagenesis. However, two Tn5 mutants were outside the region included in pBF101 (about 35 kb on the map). These mutants could exert a polar effect on the genes required for pilus synthesis or affect genes of sites required for transcription of the pilus genes from pED208.

From the transposon studies and cloning experiments, it is apparent that only one region of pED208 is required for pilus synthesis. Some plasmids, such as RP4 [Inc P (1)], require three distinct, separated regions for pilus production.

4. Complementation of adjacent HindIII fragments resulting in transfer.

Since pBF101 was unable to promote plasmid transfer, complementation studies were performed. Adjacent cloned HindIII fragments individually present on the compatible vectors pBR322 and pACYC184 were cotransformed into AA125, and transfer proficiency was tested. The results are presented in Table II.2. Bacteria carrying both pBF101 and pBF105 were able to conjugate, but only pBF105 was transferred. This suggests that the 23-kb fragment must contain the origin of transfer (*oriT*). Whether or not pBF101 contains all of the transfer genes except *oriT* is not known, since pBF105 may be supplying additional gene products.

TABLE II.1

Size and Location of Cloned pED208 Fragments

Plasmid	Vector	Fragment (Size [Kb])	Position
pBF101	pACYC184	HindIII (27.5)	7.5-34
pBF102	pACYC184	HindIII (23)	34-57
pBF103	pACYC184	HindIII (20.5)	57-77.5
pBF104	pACYC184	HindIII (12)	79-1
pBF105	pBR322	HindIII (23)	34-57
pBF106	pBR322	HindIII/XhoI (12)	22-34
pBF111	pBR322	XhoI (7.5)	14.5-22

TABLE II.2

Phage Sensitivity and Transfer Ability of Various pED208 Clones

Plasmid	f1 Sensitivity	Transfer
pBF101	+	-
pBF102	-	-
pBF103	-	-
pBF104	-	-
pBF105	-	-
pBF106	-	-
pBF111	-	-
pBF101/pBF105	+	+
pBF102/pBF106	-	-
pBF102/pBF111	-	-

Figure II.1

Agarose gel (0.6%) electrophoresis of various restriction enconuclease digests of pED208. Lane 1, HindIII; land 2, HindIII/PvuI; lanes 3 and 11, PvuI; lanes 4 and 10, lambda/HindIII plus lambda/EcoRI; lane 5, BamHI/PvuI; lane 6, BamHI; lane 7, BamHI/XhoI; lane 8, XhoI; lane 9, PvuI/XhoI; lane 12, SstI/PvuI; lane 13, SstI.

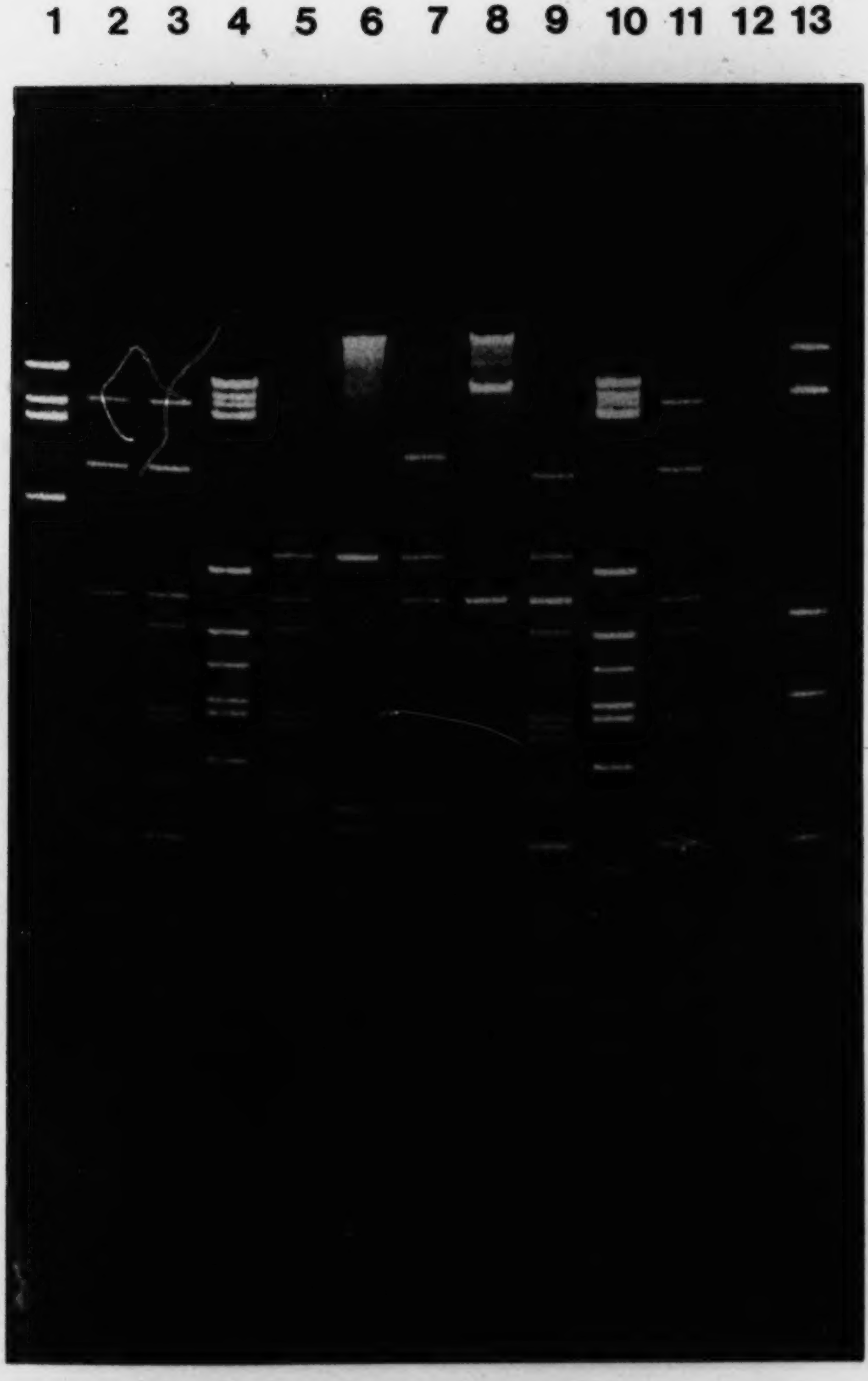

1
2
3
4
5
6
7
8
9
10
11
12
13

Figure II.2

Restriction endonuclease map of pED208. The numbered divisions represent kb.

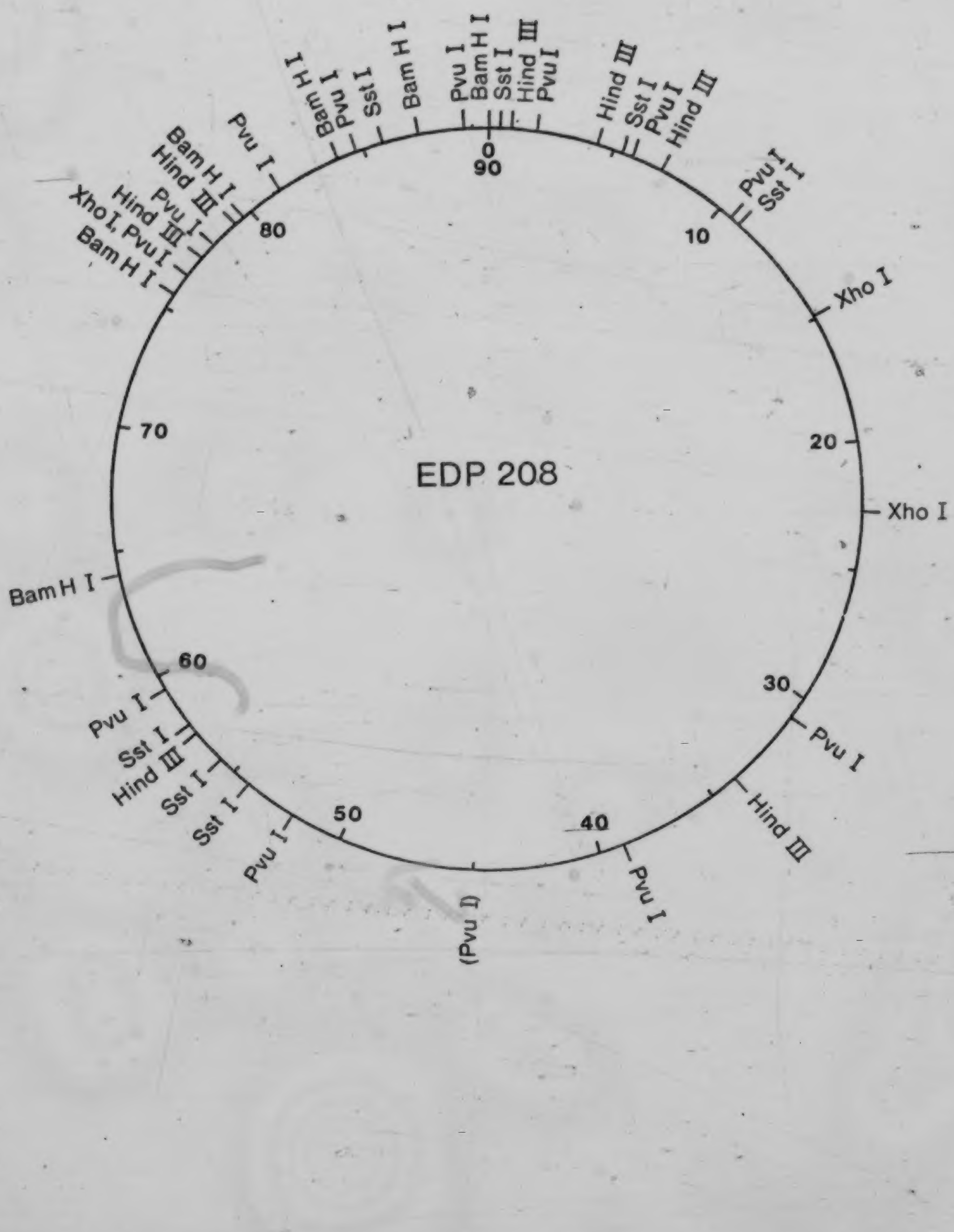
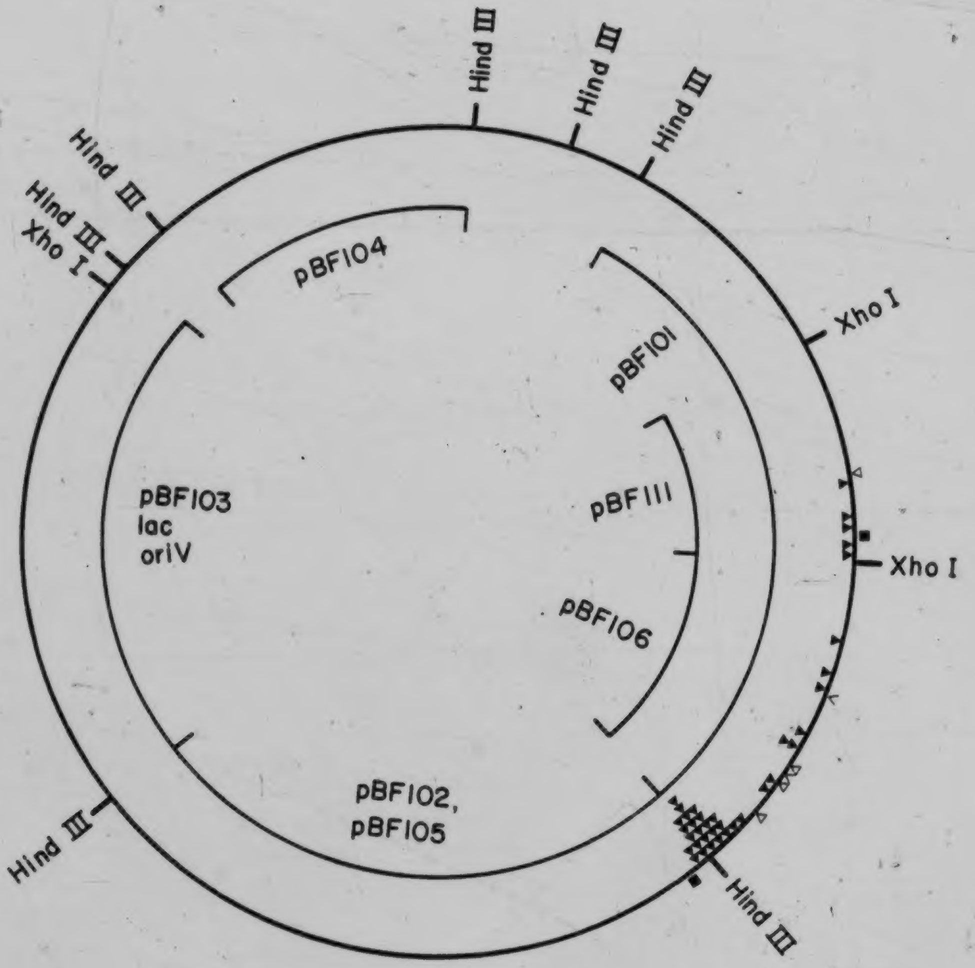

EDP 208
0
90
Pvu I
Bam H I
Sst I
Hind III
Pvu I
Hind III
Sst I
Pvu I
Hind III
Pvu I
Sst I
10
Xho I
20
Xho I
30
Pvu I
Hind III
40
Pvu I
(Pvu I)
50
Pvu I
Sst I
Sst I
Hind III
Sst I
Pvu I
60
Bam H I
70
Bam H I
Xho I, Pvu I
Hind III
Pvu I
Hind III
Bam H I
Pvu I
80
Bam H I
Pvu I
Sst I
Bam H I

Figure II.3

*Hin*dIII/*Xho*I map of pED208 showing the various clones that were constructed (Table II.1). Transposon insertions that block pilus production are also shown. Tn5 insertions are represents by ▼. Tn1 insertions are represented by ▽. The mutant with two Tn1's is marked by V, and the Tn1 insertion/deletion mutant is marked by two ■'s, with the deletion occurring between.


Hind III
Hind III
Hind III
Hind III
Hind III
Xho I
Xho I
Xho I
Hind III
Hind III
pBF104
pBF101
pBF111
pBF106
pBF103
lac
oriV
pBF102,
pBF105

D. BIBLIOGRAPHY

1. **Barth, P.T.** 1979. RP4 and R300B as wide host-range plasmid cloning vehicles, p. 399-410. In K.N. Timmis and A. Puhler (ed.), Plasmids of medical, environmental and commercial importance. Elsevier/North-Hollonad Publishing Co., Amsterdam.

2. **Birnboim, H.C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1523.

3. **Bradley, D.E.** 1981. Conjugative pili of plasmids in Escherichia coli K-12 and Pseudomonas species. p217. In S.B. Levy, R.C. Clowes, and E.L. Koening (ed.), Molecular biology, pathogenicity, and ecology of bacterial plasmids. Plenum Publishing Corp., New York.

4. **Bradley, D.E., J.N. Coetzee, T. Bothma, and R.W. Hedges.** 1981. Phage Folac: an Folac plasmid-dependent bacteriophage. J. Gen. Microbiol. 126:405-411.

5. **Chang, C.Y., and S.N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. 134:1141-1156.

6. **Crosa, J.H., and S. Falkow.** 1981. Plasmids, p. 266. In P. Gerhardt, R.G.E. Murray, R.N. Cosilow, E.W. Nester, W.A. Wood, N.R. Krieg, and G.B. Phillips (ed.), Manual of methods for general bacteriology. American Society for

Microbiology, Washington, D.C.

7. **Davis, R.W., D. Votstein, and J.R. Roth.** 1980. A manual for advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Sping Harbor, New York.

8. **Delucia, P. and J. Cairns.** 1969. Isolation of an E. coli strain with a mutation affecting DNA polymerase. Nature (London) **224**:1164-1167.

9. **Falkow, S., and L.S. Baron.** 1962. Epsisomic element in a strain of Salmonella typhosa. J. Bacteriol. **84**:581-589.

10. **Heffron, F., P. Bedinger, J.J. Champoux, and S. Falkow.** 1977. Deletions affecting the transposition of an antibiotic resistance gene. Proc. Natl, Acad. Sci. U.S.A. **74**:702-706.

11. **Heffron, F., B. McCarthy, H. Ohtsubo, and E. Ohtsubo.** 1979. DNA sequence analysis of the transposon Tn3: three genes and three sites involved in transposition of Tn3. Cell **18**:1153-1163.

12. **Hirschel, B.J., and D.E. Berg.** 1982. A derivative of Tn5 with direct terminal repeats can transpose. J. Mol. Biol. **155**:105-120.

13. **Meyers, J.A., D. Sanchez, L.P. Elwell, and S. Falkow.** 1976. Simple agarose gel electrophoretic method for identification and characterization of plasmid deoxyribonucleic acid. J. Bacteriol. **127**:1529-1537.

14. **Portnoy, D.A., S.L. Mosley, and S. Falkow.** 1981. Characterization of plasmids and plasmid-associated

determinants of Yersinia enterocolitica pathogenesis. Infect. Immun. 31:775-782.

15. **Robinson, M.K., P.M. Bennett, S. Falkow, and H.M. Dodd.** 1980. Isolation of a temperature sensitive derivative of RP1. Plasmid 3:343-347.

16. **Rothstein, S.J., R.A. Jogensen, J.C.P. Yin, Z. Yong-Di, R.C. Johnson, and W.S. Reznikoff.** 1980. Genetic organization of Tn5. Cold Spring Harbor Symp. Quant. Biol. **45**:99-105.

17. **Smith, H.R., G.A. Willshaw, and B. Rowe.** 1982. Mapping of a plasmid, coding for colonization factor antigen I and heat-stable enterotoxin production, isolated from an enterotoxin strain of Escherichia coli. J. Bacteriol. **149**:264-275.

18. **Sutcliffe, J.G.** 1978. pBR322 restriciton map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. Nucleic Acids Res. **5**:2721-2728.

19. **Willetts, N., and R. Skurray.** 1980. The conjugation system of F-like plasmid. Annu. Rev. Genet. **14**:41-76.

Chapter III

LOCALIZATION, CLONING, AND SEQUENCE DETERMINATION OF THE CONJUGATIVE PLASMID COLB2 PILIN GENE[1]

A. INTRODUCTION

Conjugative pili have long been known to be an essential structure for conjugal DNA transfer. Certain bacteriophages, such as f1 and R17, also require a pilus for infection. The paradigm and probably the most extensively studied conjugation system is F (for a review, see reference 23). F belongs to the incompatibility group IncFI and is a 100 kb plasmid. ColB2 has a related, but not identical, conjugation system. ColB2 is a colicin-producing plasmid (6) belonging to the IncFII incompatibility group. It has a type II pilus (as opposed to F's type Ia), and has many other transfer functions similar to F (23).

Since F makes relatively few pili (one or two/cell), and ColB2 produces 5-10 pili/cell (3), biochemical studies on the ColB2 pilus protein have been more productive than those on F. In this paper we have cloned and sequenced the

[1] A version of this chapter has been published. Finlay, B.B., L.S. Frost, and W. Paranchych. 1984. J. Bacteriol. 160:402-407.

DNA that encodes the ColB2 pilin protein, and compared this sequence to the recently sequenced F pilin (10). Protein chemistry was used to confirm the ColB2 pilin sequence, to identify the putative signal cleavage location and to identify the blocked amino terminus. The restriction map of the entire ColB2 plasmid is presented, as is the location of the ColB2 transfer operon, and the ColB2 pilin gene. DNA homology studies between F and ColB2 are also presented.

B. MATERIALS AND METHODS

1. Bacterial Strains and Plasmids.

Derepressed ColB2fdr (12) plasmid was carried in the host strain JC6256 (Escherichia coli K12 F⁻ trp lac) and was kindly provided by N.S. Willetts (University of Edinburgh, Scotland). pBR322 containing chimeras were transformed into AA125 (his trp bio Str^r). JM83 (22) was the organism used for transformation with pUC8, and JM103 was used for M13 cloning and sequencing (10). pED208 was carried in the host strain JC6256 (8). pED851 is a chimera consisting of the large BamHI fragment of F (containing the entire F tra operon) cloned into pBR322 (14).

2. Isolation of plasmid DNA.

The cleared lysate method described by Crosa and

Falkow (7), followed by CsCl density centrifugation was used to obtain most DNA in large quantities. Small quantities were isolated by the method of Birnboim and Doly (4).

3. Recombinant DNA Techniques.

These are described in Finlay et al. (8). [^{32}P]-labelled DNA was prepared by labelling DNA *in vitro* with [alpha-^{32}P]dATP (New England Nuclear Corp.) by nick translation as described by Maniatis et al. (15) to a specific activity of 10^7 to 10^8 cpm/microgram of DNA.

4. DNA Filter Hybridization.

After gel electrophoresis, DNA was transfered to nitrocellulose filters by the method of Southern (21). Hybridizations were done at 37 degrees with 50% formamide as described by Portnoy et al. (19).

5. Colony Filter Hybridization.

This was done as described in (16).

6. Chimeras Using pUC8.

After ligation and transformation into JM83, transformants were plated onto plates containing 5-bromo-4-chloro-3-indoxyl-beta-D-galactoside (BCIG) and isopropyl-beta-D-thio-galactoside (IPTG) as described in

(22). White colonies represented inserts.

7. M13 Dideoxy Sequencing.

These methods are described in (10).

8. ColB2 Pilin Purification.

ColB2 pili were purified from ED2601, a fla$^-$ mutant of JC3272 (1) carrying the ColB2 (IncFII) plasmid (6). F and ColB2 pili were purified using a method described previously (2). The F and ColB2 pilin subunits were prepared by treating purified pili with 1% SDS and precipitating the pilin protein with acetone to remove excess detergent (9).

9. Enzymatic Digestion of Pilin.

Digestion of ColB2 pilin (500 nmol) with TPCK-trypsin (enzyme-substrate ratio: 1/50) was performed as described previously (9). The resultant tryptic peptides were fractionated on a Biogel P4 column using 0.05M NH_4HCO_3, pH 8.0 as a buffer and were further purified by high voltage paper electrophoresis at pH 1.8 (9).

10. Tryptic Peptide Sequencing.

Tryptic peptides of ColB2 pilin were sequenced manually using the Edman degradation procedure outlined by Gray (11). Amino terminal residues were identified by dansylation and chromatography on polyamide sheets (17).

11. Complementation Assay.

Complementation assays were performed using AA125 carrying the plasmid JCFL1, which contains a point mutant in the traA gene of the wild type F plasmid, blocking F pilus synthesis and F transfer during bacterial conjugation (1). The recipient strain was ED24 (F^- lac Spc^r). Qualitative complementation was measured by the replica-plate mating method of Clark and Margulies (5).

C. RESULTS

1. Restriction Enzyme Analysis of ColB2.

Isolated ColB2 plasmid DNA was digested with five restriction enzymes-BamHI, EcoRI, HindIII, SalI, and XhoI. Digestions were performed with either one or two of these enzymes present in all possible combinations. The restriction fragments were analysed by separation on a 0.6% tris-acetate agarose gel, followed by ethidium bromide staining. The fragments were sized by comparing mobilities with lambda/HindIII DNA. Summation of all restriction fragments suggested that the ColB2 plasmid is approximately 96 kilobases in size (63 x 10^6 daltons). This is slightly lower than the previously reported value of 70 x 10^6 daltons (106 kb) that was obtained by contour length measurement by electron microscopy (6). However, this is similar to the F plasmid (100 Kb) and other large conjugative plasmids such as pED208 (90 Kb) (8).

By performing digestions with single and combinations of restriction endonucleases, a restriction map of the ColB2 plasmid was generated (Fig. III.1). A striking feature of the ColB2 restriction map is the asymmetric distribution of restriction sites. This is not unusual for large conjugative plasmids, such as pED208 (8). There is usually a dearth of restriction sites in the transfer operon region, while other regions usually contain a high density of sites, possibly suggesting that restriction sites within transfer operons are undesirable (23).

2. DNA Homology with the F tra Region.

Since F and ColB2 are known to have similar transfer operons (23), DNA homology, at least in the transfer operon was expected. When Southern blots of F DNA digested with BamHI or HindIII were probed with ^{32}P-labelled ColB2 DNA, a single band in each digest was found to be homologous and in each case this band was the largest restriction fragment (data not shown). This region of F (which corresponds to the large HindIII and BamHI fragments) includes the entire F transfer operon, and a small amount of DNA which is not transfer operon associated (23), suggesting that the only homology between these two plasmids occurs in their transfer operons. The large BamHI fragment of F has been cloned into pBR322 (14), generating a recombinant with an intact F transfer operon. When this chimera (pED851) was digested with EcoRI and probed with ColB2 DNA, all of the major EcoRI fragments of F (23), were homologous to ColB2 (Fig. III.2, D and E). Three small EcoRI fragments of F were not detected on a 0.6% agarose gel and thus no homology comparisons for these fragments can be made.

Also included as a control in this experiment was the conjugative plasmid pED208. This 90 Kb plasmid has a transfer operon similar in size to F (8). However, this plasmid showed no homology with ColB2 DNA, even though f1 phage can infect cells carrying either plasmid. Their pili are also unrelated with respect to their amino acid composition (3), suggesting analogous but not homologous transfer operons.

When the reciprocal blots were tested (i.e., ^{32}P-labelled pED851 was used to probe various digests of ColB2), homology was found with EcoR1 fragments of 37 Kb (map positions 22 to 59) and 9 Kb (map positions 59-68), and Sal I fragments of 27 Kb (map positions 27 to 54) and 10.5 Kb (map positions 54.5 to 65) (Fig. III.2, B, C). This suggested that the major portion of the ColB2 transfer region is between map positions 27 and 65. Since weak homology was also detected with a 28 Kb SalI fragment (map positions 65 to 93) and a 20 Kb EcoRI fragment (map positions 68 to 88), weak homology with F also occurs between map positions 68 to 88. It is not clear whether this weak homology involves DNA which is within or immediately adjacent to the transfer region.

Although the homology between the F and ColB2 plasmids is rather extensive in the transfer region, it is not identical. The F transfer region contains several EcoRI sites (23), while Col B2 has only one (or a few clustered) EcoRI site(s) in its transfer operon.

Since our interest was in localizing the ColB2 pilin gene, various digests of ColB2 DNA were probed with a chimera that contained traA, the pilin gene of F (pED891) (10). This chimera consists of a 1.4 kb PstI fragment that contains traA, traL and traE, ligated into pBR328 (10). When Southern blots of ColB2 DNA were hybridized with this chimera, a 10.5 kb SalI fragment of ColB2 was found to be homologous to this F fragment (see Fig. III.2 A, B). This SalI fragment is the smallest of the three SalI fragments of ColB2 that were found to be homologous to the F transfer operon. Correspondingly, the overlapping 9 kb EcoRI fragment was found to be homologous to the F traALE probe, as was the 6 kb EcoRI/SalI fragment which contained 6 kb of DNA common to both the EcoRI and SalI fragments. This localized the ColB2 pilin gene to this 6 kb region from map positions 59 kb to 65 kb (Fig. III.1).

If the genetic ordering of the transfer operon of ColB2 and F are similar, one would expect the operon to be transcribed beginning near the traA gene (in the EcoRI/SalI 6 kb fragment) and proceeding into the large (27 kb) SalI fragment. This operon orientation allows alignment of the EcoRI sites to those within the F operon flanking traV (f15) (23), and the SalI sites within traM and traC. The EcoRI sites of f17 and f19 are not present in ColB2. These restriction sites fall within traG and the surface exclusion genes. Since ColB2 and F have different surface exclusion systems (23), differences in DNA in this region are not surprising.

3. Cloning of traA of ColB2.

Since a very specific pilin gene probe was available (F traALE, containing the F pilin gene), we used colony blots to detect ColB2 pilin gene clones (16). PstI digestions of ColB2 plasmid DNA were ligated into a positive insertional vector, pUC8 (22). Potential chimaeras were patched onto nitrocellulose, lysed, baked, and probed with the 1.4 kb labelled PstI fragment from the F traALE chimera (see Materials and Methods). Nine colonies out of 400 were positive for hybridization with the F traALE probe. Upon digestion of the chimeras, all were found to contain a 1.4 kb PstI insert, the same size as the homologous fragment in F.

To prove that the pilin gene was contained within this fragment, we tested the chimera's ability to complement an amber mutant in the F pilin gene (F traA1), while all other F transfer operon proteins are intact and functional. Since ColB2 and F are known to have very similar pili (as judged by R17 phage sensitivity, immunological cross-reactivity, and amino acid analysis [23]), complementation should occur.

These chimeras were transformed into E. coli AA125 carrying the JCFL1 (traA1) plasmid and their ability to complement the traA1 point mutant and restore transfer ability to the F plasmid was compared to that for an equivalent chimera containing the 1.4 kb Pst I fragment from F (Table III.1).

Both chimeras complemented the mutation equally well at a level 1000-fold less than transfer ability of the wild type F plasmid (see Table III.1). Chimeras containing the 1.4 kb PstI fragment in the opposite orientation failed to complement at all, indicating that transcription of the pilus gene was under control of the lac promoter supplied in the vector pUC8.

In addition, these chimeras restored sensitivity to two pilus specific phages, f1 and R17, for cells carrying JCFL1 (traA1). These results indicate that the genes for F and ColB2 pilin can be used interchangeably to supply pilin for the tra proteins of the F transfer region to synthesize an F-like pilus. The chimera containing the ColB2 pilin gene on a 1.4 kb PstI fragment inserted in pUC8 in the correct orientation was called pBF203.

4. Sequencing of the ColB2 Pilin Gene.

All sequencing was done by Sanger's M13 dideoxy method (20). The entire 1.4 kb fragment was cloned into M13mp8/*Pst*I and sequenced from both ends (Fig. III.3). The internal *Sau*3a 363 base fragment (which contains the C-terminal end of *traA*) and the 800 base *Sau*3a/PstI fragment (which contains *traL* and *traE*) were also sequenced. The sequence for ColB2 pilin (Fig. III.3) is virtually identical to that for F pilin (10) except for the deletion of six base pairs in the region defining the N-terminal of processed pilin. Thus the sequence GGCAGCAGT for the F pilin gene is shortened to CAG in ColB2. There were no differences in the 800 bp *Sau*3a/*Pst*I fragment and therefore the *traL* and *traE* proteins of F and ColB2 are identical.

5. Sequence of Tryptic Peptides From ColB2 Pilin.

To ensure that the sequence of ColB2 (and F) pilin were correct, ColB2 pilin was digested with TPCK-treated trypsin. The soluble peptides were separated on a P4 column at pH 8.0, followed by paper electrophoresis at pH 1.8 and elution of the peptides from the paper. Three soluble peptides were purified and subjected to amino acid compositional analysis (Table III.2). The peptides (CT1 to

CT3) were analyzed by dansylation to identify the N-terminal amino acid residues. The peptides were sequenced as far as possible using manual Edman degradation and the derivatized amino acids were identified by thin layer chromatography on polyamide sheets (17). Sequence information was obtained on two of the peptides (CT2, CT3) and this was aligned with the sequence of ColB2 pilin derived from the DNA (residues 67-77) (Table III.2, Fig. III.3). The third peptide, CT1, had a blocked N-terminus and could not be sequenced. However, there is good correlation between the amino acid composition of this peptide and the peptide in ColB2 pilin from residues 52-66. The sequences of the ColB2 pilin peptides agrees with that predicted by the DNA sequence of ColB2 pilus gene and represents the first 25 residues in the N-terminal portion of the protein. The remaining 43 amino acids in the C-terminal protion of the protein as predicted by the DNA sequence are very hydrophobic in nature and it is not surprising that these peptides from ColB2 pilin were insoluble and not readily isolated.

6. The N and C-Terminals of ColB2 pilin.

The N- and C-termini of ColB2 pilin were identified using the procedures described by Frost et al. (10). The amino acid analysis of the blocked N-terminal peptide derived from a pronase digestion of ColB2 pilin and the amino acids released from the C-terminal after digestion with Carboxypeptidase A are given in Table III.3. A peptide which had an overall acidic charge and was not reactive with ninhydrin was isolated and characterized as containing alanine, glutamic acid (or glutamine), and glycine in a ratio of 1:1:0.33. This was consistent with an amino terminal starting at ala_{52} where the ratio of dipeptide (ala-gln) was twice that of the tripeptide (ala-gln-gly). Proton NMR spectroscopy (data not shown) showed that the alanine was probably blocked with an acetyl group where the ratio of acetyl group protons to the protons on the methyl group (beta carbon) on alanine are 1:1.

This is reminiscent of the acetylated N-terminal alanine of F pilin (10). However the peak for the acetyl group protons (2.0 ppm) is shifted 0.05 ppm downfield for F pilin, reflecting the difference in the amino acid subsequent to the alanine. Thus the N-terminal for ColB2 pilin is acetyl-ala-gln-gly-gln- and that for F pilin is acetyl-ala-gly-ser-ser-gly-gln-. This difference is reflected both in the DNA sequence and the position of the protons of the N-terminal amino acid residue in an NMR spectrum.

The results of the Carboxypeptidase A digestion of ColB2 are shown in Table III.3. The amino acids released were leu:gly:val:ala in a ratio of 1.0:0.81:1.52:0.8 which corroborate the sequence at the C-terminal predicted by the

DNA sequence of ColB2 pilin.

D.DISCUSSION

Sequence comparisons between F and ColB2 DNA encoding the _traA_ gene revealed only a few base changes. Six nucleotide changes are found in the region encoding the signal peptide of pilin, all of which are conservative. These differences, and the nucleotide differences found in the region encoding the N-terminal of pilin, are the only nucleotide differences found in the _traA_ gene (and in the entire _traALE_ fragment) of F and ColB2. The significance of these conservative changes in the region encoding the signal peptide is not known.

The only region of the _traA_ gene which affects the pilin protein sequence is a six base pair deletion in the region defining the N-terminal of processed pilin. The significance of these changes at the DNA level clustered at the N-terminal of the pilin protein are not immediately obvious. However, these changes do generate the only alterations in amino acid content between the two types of pilin and this is probably the basis of the serological differences between F and ColB2 pili (reported by Orskov & Ketyi [18]). These authors reported much cross-reactivity immunologically between F and ColB2, but, "in addition, (ColB2) had its own specific factor" (18).

On the basis of their serology and the efficiencies of plating of several pilus-specific phages, F and ColB2 pili have been subdivided into type Ia and II pili, respectively (23) and this difference could be explained by the change in the N-terminals of the two pili. Recently, R. Hodges et al. (13) have shown that substituting different amino acids at the various positions after the N-acetyl threonine in a peptide (N-terminal dodecapeptide) that constitutes the major antigenic determinant of pED208 pilin greatly alters the antigenicity of the peptide. Thus it is not unreasonable to expect that, while F and ColB2 pilin share an acetylated alanine at their N-terminals, the change in the sequence from (F)-gly-ser-ser-gly- to (ColB2) -gln-gly- could account for their serological differences.

Two synthetic peptides representing the first 8 amino acids of F pilin (Ac-ala-gly-ser-ser-gly-gln-asp-leu-COOH) and the first 6 amino acids of ColB2 pilin (Ac-ala-gln-gly-gln-asp-leu-COOH) were synthesized using the protocol of Worobec et al. (24) and tested for their ability to compete with pure intact pili for anti-pilus antibodies using a competition ELISA procedure. The results (to be reported in detail elsewhere) showed that the F synthetic peptide competes very strongly for polyclonal anti-F pilus antibodies while the ColB2 peptide competed weakly. These observations confirm that the antigenic differences between F and ColB2 pili reside in the N-terminal alterations in the protein sequence and that the blocking group is an acetyl group.

A major difference between F and ColB2 pili is the inability of ColB2 pili to retract in the presence of cyanide and a 20-fold reduction in sensitivity to filamentous phages such as f1 (Moore, Paranchych and Willetts, unpublished results). It is unlikely that these differences can be accounted for by the minor change in the sequence at the amino-termini of these pili. It seems more likely that an alteration in the assembly-disassembly mechanism in the membrane of F or ColB2-containing cells could explain all the phenomena noted above. In F, the traYZ operon is under control of traJ and transcription begins at the promoter at the base of the operon (23). There could be differences in the organization and operation of the equivalent operon in ColB2 that leads to a higher number of pili per cell. We are currently studying the traJ/traYZ system in ColB2 in order to identify the differences in operon expression as the basis for the phenotypic variatons between F and ColB2 pilus systems.

In conclusion, the sequence for ColB2 pilin and F pilin is identical, except at the amino terminal, where 2 codons are deleted or altered. Thus the F N-terminal is Ac-ala-gly-ser-ser-gly-gln-asp and ColB2 is Ac-ala-gln-gly-gln-asp-. The only other nucleotide changes are conservative, and occur within the signal peptide.

Their transfer operons show extensive homology in hybridization studies. Several *EcoRI* site differences occur near the surface exclusion system region, but both *EcoRI* and *SalI* sites are conserved within the pilus producing region. *TraL* and *traE* appear to have identical sequences in ColB2 and F and the region upstream of *traA* is identical in both. Thus, the only difference in the *traALE* fragment of the F and ColB2 plasmids is the six conservative nucleotide changes in the leader sequence of *traA* and the six base pair deletion and two conservative changes in the amino terminal of *traA*.

TABLE III.1

Complementation Ability of Chimeric Plasmids Containing F or ColB2 Pilin Gene.

Plasmid	Transconjugates per 10^6 donor cells^a	Complementation Ability (%)	Comments and References
JCFLO	27	100	1
pUC8	0.0 ($<10^{-6}$)	0	23
pLF16	0.025	0.09	10; F*traALE* in pUC8
pBF203	0.027	0.10	ColB2 *traALE* in pUC8
pBF209	0.0 ($,10^{-6}$)	0	ColB2 *traALE* in pUC8, reverse oreintation

a. The number of donor cells [AA125/JCFL1 (*traA*1) + pUC8 chimera] was determined by using MacConkey lactose plates containing ampicillan (50 micrograms/ml). The number of transconjugants from the mating with the recipient strain ED24 was measured by using minimal lactose plates containing spectinomycin (100 micrograms/ml). JCFLO is the wild type F plasmid used in place of JCFL1 as the positive control.

b. Complementation ability is expressed as a percent value, where the wild-type F plasmid JCFLO was set at 100%.

TABLE III.2

Amino Acid Composition of Tryptic Peptides From ColB2 Pilin.

Peptide Amino Acid	nmol (ratio) of[a]: CT1	CT2	CT3
Asx	37.0 (2.13)	0 (0)	16.6 (1.1)
Thr	28.5 (1.64)	12.8 (1.0)	0 (0)
Ser	23.0 (1.32)	0 (0)	25.7 (1.7)
Glx	34.1 (1.96)	0 (0)	0 (0)
Gly	37.4 (2.15)	23.3 (1.8)	0 (0)
Ala	34.8 (2.0)	14.3 (1.1)	0 (0)
Val	10.4 (0.6)	0 (0)	13.6 (0.9)
Met	15.8 (0.91)	0 (0)	0 (0)
Ile	0 (0)	0 (0)	0 (0)
Leu	18.3 (1.05)	0 (0)	0 (0)
Tyr	0 (0)	0 (0)	0 (0)
Phe	0 (0)	13.0 (1.0)	0 (0)
Lys	17.4 (0.85)	13.1 (1.0)	15.1 (1.0)

a. The conditions for tryptic digestion are givien in the text. Approximately 20 nmol of each peptide was hydrolyzed for 18 h at 110°C in constantly boiling 6 N HCl in vacuo. Amino acids in amounts of 5 nmol or less are recorded as zero. The sequence of the peptides as determined by manual Edman degradation was CT2, -Ala-Thr-Phe-Gly-Lys; and CT3, -Asp-Ser-Ser-Val-Val-Lys. The residue numbers were CT1, 52 to 66; CT2, 67-71; and CT3, 72 to 77.

TABLE III.3

Analysis of Amino Acids at the N Terminal and C terminal of ColB2 Pilin.

Amino Acid	nmol (ratio) at: N terminal[a]	C terminal
Asx	15.8 (0.26)	
Thr	6.2 (0.1)	
Ser	7.6 (0.12)	
Glx	60.9 (1.0)	
Ala	56.3 (0.92)	22.5 (0.8)
Gly	20.1 (0.33)	22.7 (0.8)
Val		42.6 (1.52)
Leu	4.3 (0.07)	28.0 (1.0)
Met		9.4 (0.34)

a. The conditions for proteolytic digestion are given in the text. The analyzis of amino acids released from the C terminal represents that expected from 50 nmol of ColB2 pilin. Amino acids present in amounts os 5 nmol or less are not reported. The predicted amino acid sequence is Nterminus, N-acetyl-Ala-Gln-Gly-Gln-Asp-Leu; and C terminus, Met-Ala-Val-Val-Gly-Leu.

Figure III.1

Restriction map of ColB2 showing the locations of the transfer operon and the pilin gene (traA) of ColB2. Divisions are in Kb.

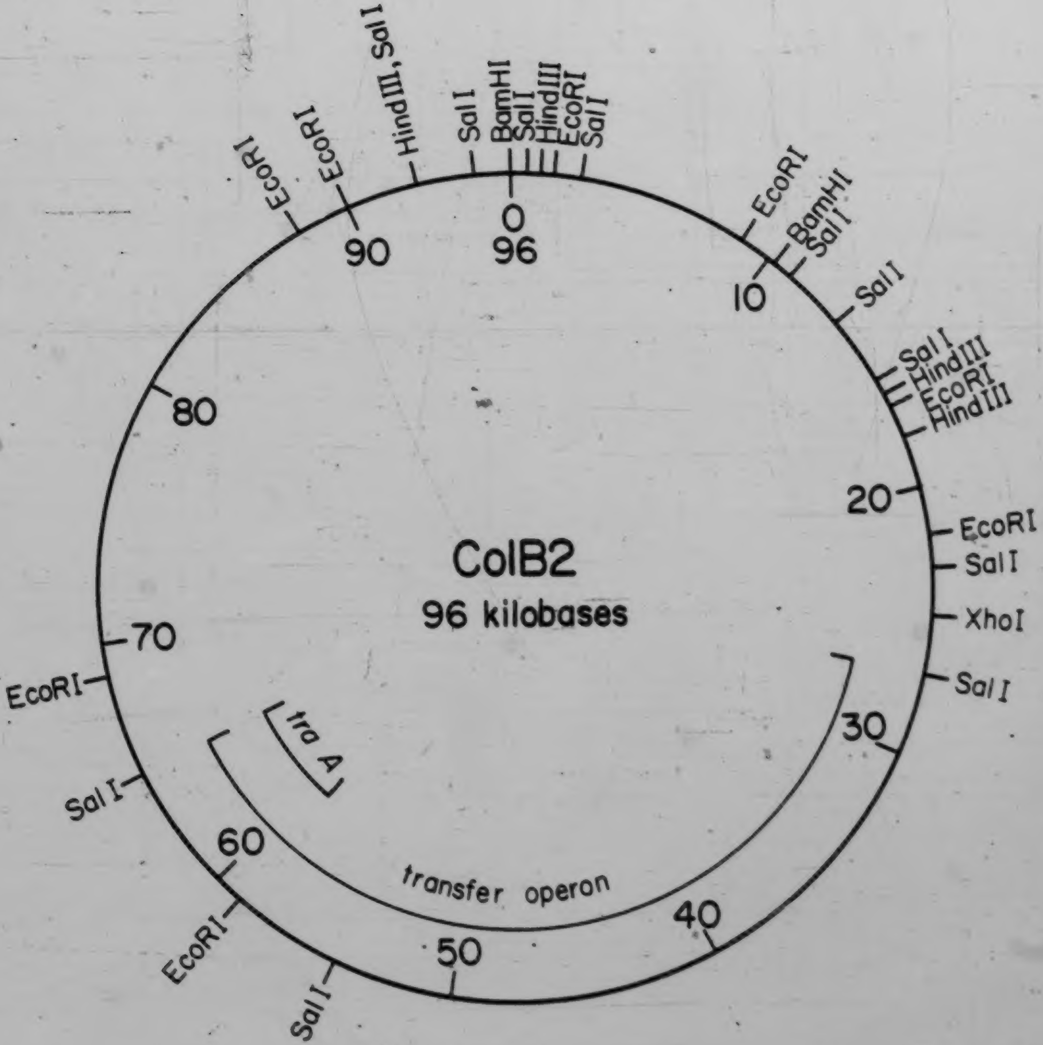

ColB2
96 kilobases
0
96
10
20
30
40
50
60
70
80
90
EcoRI
EcoRI
HindIII, SalI
SalI
BamHI
SalI
HindIII
EcoRI
SalI
EcoRI
BamHI
SalI
SalI
SalI
HindIII
EcoRI
HindIII
EcoRI
SalI
XhoI
SalI
SalI
EcoRI
SalI
EcoRI
transfer operon
tra A

Figure III.2

Southern blots showing homology between ColB2 and F plasmids. (A) Autoradiograph of Southern blot of B, hybridized with F *traALE* cloned in pBR328; (B) 0.6% agarose gel of various ColB2 digests; (C) autoradiograph of Southern blot of B, probed with the entire F transfer operon cloned in pBR328 (pED851). For A, B, and C, lanes: 1, lambda/*Hin*dIII; 2, ColB2-*Eco*RI; 3, ColB2-*Eco*RI/*Sal*I; 4, ColB2-*Sal*I. (D). Autoradiograph of Southern blot of E, probed with ColB2 DNA; (E), 0.6% agarose gel of Southern blot of various digests of the transfer operon of F cloned into pBR322 (pED851). For D and E, lanes: 1, pED851-*Eco*RI; 2, pED851-*Bam*HI; 3, pED851-*Bam*HI/*Eco*RI; 4, pED208-*Hin*dIII; 5, lambda-*Hin*dIII.

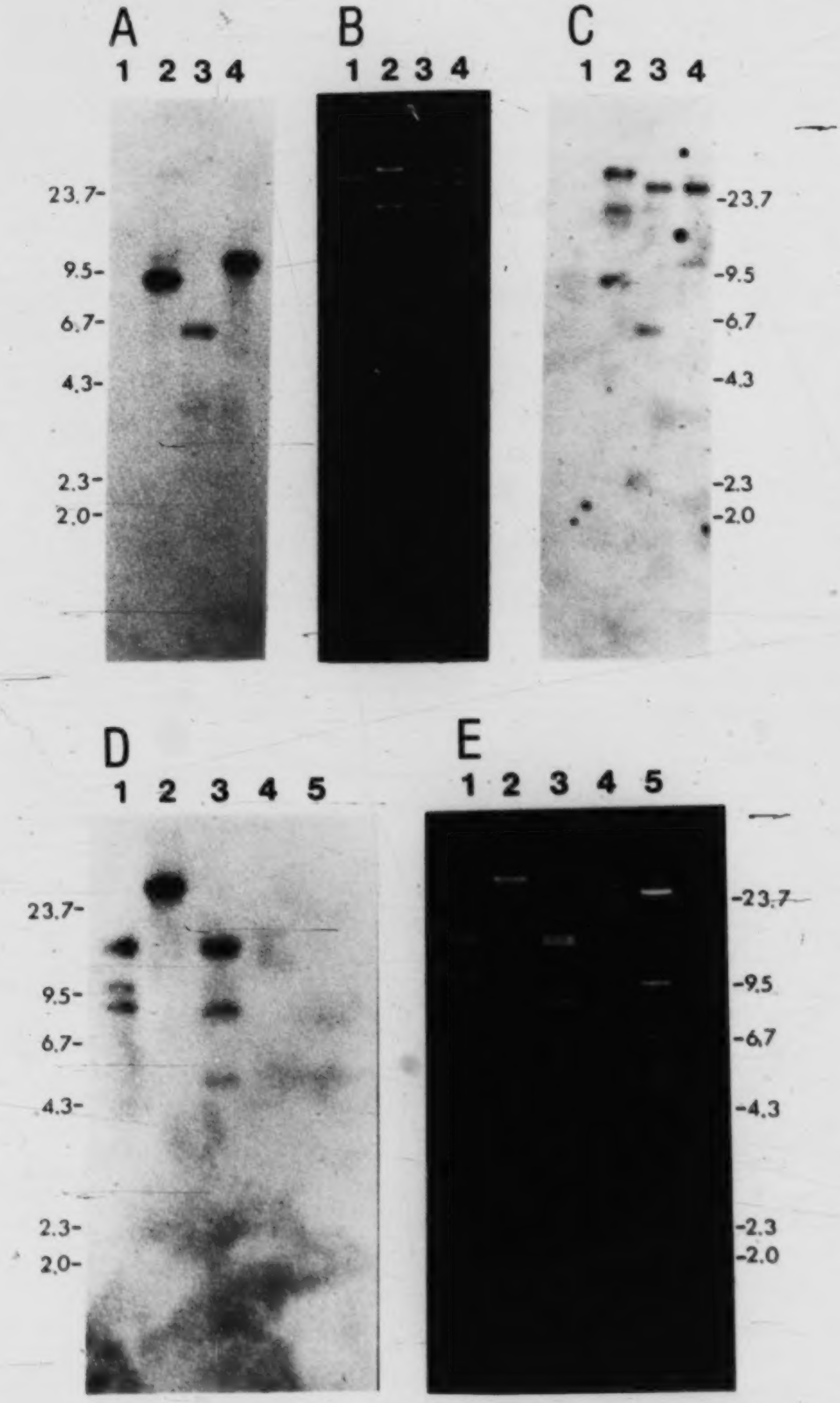

A
1 2 3 4
23.7-
9.5-
6.7-
4.3-
2.3-
2.0-
B
1 2 3 4
C
1 2 3 4
-23.7
-9.5
-6.7
-4.3
-2.3
-2.0
D
1 2 3 4 5
23.7-
9.5-
6.7-
4.3-
2.3-
2.0-
E
1 2 3 4 5
-23.7
-9.5
-6.7
-4.3
-2.3
-2.0

Figure III.3

Entire nucleotide sequence of the ColB2 pilin gene. The proposed amino acid sequence of ColB2 propilin is shown in line 1. The DNA sequence of the antisense strand is shown in line 2. In line 3 are the nucleotide changes found in F. The insert into the ColB2 sequence at amino acid 53 indicates the additional two codons found in the F DNA sequence. The putative signal peptidase cleavage site is represented by the vertical arrow.

```
                  5                  10                  15                  20                  25
MET ASN ALA VAL LEU SER VAL GLN GLY ALA SER ALA PRO VAL LYS LYS LYS SER PHE PHE SER LYS PHE THR ARG
ATG AAT GCT GTT TTA AGT GTT CAG GGT GCT TCT GCG CCC GTC AAA AAG AAG TCA TTT TTT TCT AAA TTC ACT CGT
            150                                                      G       200 C

                 30                  35                  40                  45                  50
LEU ASN MET LEU ARG LEU ALA ARG ALA VAL ILE PRO ALA ALA VAL LEU MET MET PHE PHE PRO GLN LEU ALA MET
CTG AAT ATG CTT CGC CTG GCT CGC GCT GTG ATC CCG GCT GCT GTT CTG ATG ATG TTC TTC CCG CAA CTG GCA ATG
                                 A          250                                      G       G

    ↓            55                  60                  65                  70                  75
ALA ALA GLN GLY GLN ASP LEU MET ALA SER GLY ASN THR THR VAL LYS ALA THR PHE GLY LYS ASP SER SER VAL
GCA GCA CAG GGC CAG GAC CTG ATG GCA AGC GGT AAC ACC ACG GTT AAG GCG ACC TTC GGT AAG GAC TCC AGT GTT
 C    C/  \   T
   |GGC AGC AGT|
   |GLY SER SER|
                 80                  85                  90                  95                 100
VAL LYS TRP VAL VAL LEU ALA GLU VAL LEU VAL GLY ALA VAL MET TYR MET MET THR LYS ASN VAL LYS PHE LEU
GTT AAA TGG GTT GTT CTG GCT GAA GTT CTG GTC GGT GCT GTC ATG TAC ATG ATG ACC AAA AAC GTC AAG TTC CTG

                105                 110                 115                 120
ALA GLY PHE ALA ILE ILE SER VAL PHE ILE ALA VAL VAL MET ALA VAL VAL GLY LEU ***
GCC GGT TTT GCC ATC ATC TCT GTA TTT ATT GCT GTG GTT ATG GCC GTC GTT GGC CTC TGA
```

E. BIBLIOGRAPHY

1. **Achtman, M., N.S. Willetts, and A.J. Clark**. 1972. Conjugational complementation analysis of transfer-deficient mutants in F*lac* in *E. coli*. J. Bacteriol. 110:831-842.

2. **Armstrong, G.D., L.S. Frost, P.A. Sastry, and W. Paranchych**. 1980. Comparative biochemical studies on F and EDP208 conjugative pili. J. Bacteriol. **141**:333-341.

3. **Armstrong, G.D., L.S. Frost, H.J. Vogel, and W. Paranchych**. 1981. Nature of the carbohydrate and phosphate associated with ColB2 and EDP208 pilin. J. Bacteriol. **145**:1167-1176.

4. **Birnboim, H.C., and J. Doly**. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1523.

5. **Clark, A.J., and A.D. Margulies**. 1965. Isolation and characterization of recombination-deficient mutants of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **53**:451-459.

6. **Clowes, R.C.** 1972. Molecular structure of bacterial plasmids. Baceriol. Rev. **36**:361-405.

7. **Crosa, J.H., and S. Falkow**. 1981. Plasmids, p. 266. *In* P. Gerhardt, R.G.E. Murray, R.N. Costilow, E.W. Nester, W.A. Wood, N.R. Krieg, and G.B. Phillips (ed.), Manual of Methods for General Bacteriology. American Society for

Microbiology, Washington, D.C.

8. **Finlay, B.B., W. Paranchych, and S. Falkow**. 1983. Characterization of conjugative plasmid EDP208. J. Bacteriol. **156**:230-235.

9. **Frost, L.S., G.D. Armstrong, B.B. Finlay, B.F.P. Edwards, and W. Paranchych**. 1983. N-terminal amino acid sequencing of EDP208 conjugative pili. J. Bacteriol. **153**:950-954.

10. **Frost, L.S., W. Paranchych, and N.S. Willetts**. 1984. The DNA sequence of F traALE which includes the gene for F pilin. J. Bacteriol. **160**:395-401.

11. **Gray, W.R**. 1967. Sequential degradation plus dansylation. Meth. Enzymol. **11**:469-475.

12. **Hausmann, C., and R.C. Clowes**. 1971. ColB2-K77, a fertility-repressed F-like sex factor. J. Bacteriol. **107**:900-906.

13. **Hodges, R.S., J.M.R. Parker, A.K. Taneja, E.A. Worobec and W. Paranchych**. 1984. Immunodominant region of EDP208 pili: Antigen-antibody interaction. Cold Spring Harbor Symp. on Quant. Biol. (In press).

14. **Johnson, D.A., and N. Willetts**. 1980. Construction and characterization of multicopy plasmids containing the entire F transfer region plasmid. Plasmid **4**:292-304.

15. **Maniatis, T., A. Jeffrey, and D.G. Kleid**. 1975. Nucleotide sequence of the rightward operator of phage lambda. Proc. Natl. Acad. Sci. U.S.A. **72**:1184-1188.

16. **Maniatis, T., E.F. Fritsch, and J. Sambrook**. 1982. p.313-314 In Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, New York.

17. **Narita, K., H. Matsuo and T. Kakajima**. 1975. End group determination. Biochem. and Biophys.8:30-103.

18. **Orskov, I., and I. Ketyi**. 1969. Studies on the antigen structure of sex fimbriae carried by a strain of Shigella flexineri 4b. Acta Bath. Microbiol. Scand. 77:249-308.

19. **Portnoy, D.A., S.L. Mosely, and S. Falkow**. 1981. Characterization of plasmids and plasmid-associated determinants of Yersinia enterocolitica pathogenesis. Infect. Immun. 31:775-782.

20. **Sanger, F., S. Nicklen, and A.R. Coulson**. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74:5463-5467.

21. **Southern, E.M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

22. **Vieira, J., and J. Messing**. 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene19:259-268.

23. **Willetts, N., and R.S. Skurray**. 1980. The conjugation system of F-like plasmids. Ann. Rev. Genet. 14:41-76.

24. **Worobec, E., A. Taneja, R.S. Hodges, and W.**

Paranchych. 1983. Localization of the major antigenic determinant of EDP208 pili at the N-terminus of the pilus protein. J. Bacteriol. 153:955-961.

Chapter IV

MAJOR ANTIGENIC DETERMINANTS OF F AND COLB2 PILI[1]

A. INTRODUCTION

Conjugative pili are filamentous surface appendages encoded by self transmissible plasmids. These structures assist in transfer of the conjugative plasmid between bacteria by initiating cell-cell contact during bacterial conjugation (1). In addition, they serve as receptors for certain pilus specific bacteriophages such as f1 (6) and R17 (8,24). These appendages consist of a repeating subunit, called pilin, arranged in a helical array (10) to give a tubular filament 8.0 nm in diameter and 1-20 micrometers in length.

F (IncFI) (5) is a 100 kb plasmid containing a 33 kb transfer region which encodes the most extensively studied conjugative pilus system (27,28). The genes required for F pilus assembly are located within the transfer (tra) operon which also encodes the gene products

[1] A version of this chapter has published.
Finlay, B.B., L.S. Frost, W. Paranchych J.M.R. Parker, and R. Hodges. 1985. J. Bacteriol. **163**:331-335.

required for DNA metabolism during transfer (28). It is believed that at least 14 proteins are involved in the F pilus assembly process. However, only two, traA and traQ, have been assigned functions thus far (20,22). TraA is the structural gene for F pilin, and encodes a 13,000 dalton protein (14,16,21) whose gene has been sequenced (12). This polypeptide is processed into a 7,000 dalton protein, possibly by traQ (22). The processed pilin molecule contains an acetylated amino terminus (12).

The ColB2 plasmid (IncFII) (5) is a 96 kilobase conjugative plasmid which encodes colicin production (13). It has a similar transfer region to that of F and encodes a closely related pilus (9). In addition, bacteria containing the ColB2 plasmid produce 5 to 10 pili per cell (2) as opposed to 1-2 pili/cell found on cells carrying the F plasmid. Whereas F pili have been classified as having a type IA pilus, ColB2 were classified as type II on the basis of their resistance to f1 phage and apparent inability to retract in the presence of cyanide (27, J. Maule and N.S. Willetts, unpublished results).

The genes encoding F and ColB2 pili were recently sequenced and it was found that F and ColB2 plasmids encode very similar proteins differing only at their amino termini (9,12). The sequence for the amino-terminus of F pilin was found to be Ac-ala-gly-ser-ser-gly-gln-asp-leu-met... while that of ColB2 pilin was

Ac-ala-gln-gly-gln-asp-leu-met... The remaining sequence of both proteins was identical.

Ketyi and Orskov (17) have observed that F and ColB2 pili are serologically distinct, but that in addition to a pilus-specific antigenic determinant or "factor", they also carried a second antigenic determinant common to both pilus types.

In the completely unrelated conjugative pilus system pED208, the acetylated amino terminal and the first five amino acids constitute the major antigenic determinant of this pilus system (29,30). Since F and ColB2 pili differ only in this region, yet have different serotypes, these N-terminal regions were probable candidates for the major antigenic determinant of F and ColB2 pilins. Studies to establish the immunological properties of F and ColB2 N-termini were thus carried out and are described in the present communication.

B. MATERIALS AND METHODS

1.Bacteria.

pED236 [derepressed ColB2fdr (ColB2) plasmid] was carried in the host strain JC6256 (Escherichia coli K-12 trp lac) which was kindly provided by N. S. Willetts, University of Edinburgh, Scotland. Bacteria containing a chimera of the entire F transfer region in pBR322 (27) were used in the purification of F pili. The background strain was JE2571 (Fim Fla Sm^r Nal^r Rif) which was provided by D.E. Bradley, Univ. of Newfoundland, St. John's, Nfld.

2. Preparation of Whole Pili.

F and ColB2 pili were purified as described in Armstrong et al. (2).

3. Antisera Production.

ColB2 and F pilus-specific polyclonal antisera were raised in New Zealand white rabbits as described by Worobec et al. (29). Both F and ColB2 antisera had an endpoint of 1×10^{-6} as judged by a direct ELISA using alkaline phosphatase-conjugated goat anti-rabbit antisera (Sigma Chemical Co. St. Louis, Mo.) and p-nitrophenyl phosphate as the substrate.

4. Antiserum Preabsorption.

One liter of *E. coli* JC6256 (4×10^8 cells/ml) grown in L-broth was centrifuged, and the cell pellet resuspended in 0.5 ml undiluted anti-ColB2 pilus antiserum. This was incubated at 37 degrees for 2 hours on a rotating platform, after which the cells were repelleted and the supernatant antiserum was stored at 4 degrees.

5. Competition ELISA.

These were performed as in Worobec et al. (30), except that microtiter wells were precoated with 2 micrograms/ml pili homologous to the antisera used. Anti-pilus antiserum was used at a dilution of 1/2000.

6. Amino Terminal Tryptic Peptide Isolation.

This was done as described in Finlay et al. (9). Approximately 2 mg of purified ColB2 pili were digested with TPCK-trypsin (E/S=1/50) to give 50 nmol of purified amino terminal tryptic peptide.

7. Cyanogen Bromide Digestion.

The peptide was dissolved in 70% formic acid and dry cyanogen bromide was added (molar ratio 300:1)(15) and the reaction was monitored by thin layer electrophoresis (11) and by the conversion of methionine to homoserine as measured by amino acid analysis.

8. Peptide synthesis and purification.

The peptides were synthesized on a Beckman synthesizer Model 990, using procedures previously described (30). All amino groups were protected at the alpha-amino position with t-butyloxycarbonyl (Boc) group and the following side chain blocking groups were used: glutamine (p-methoxybenzhydryl) and aspartic acid (o-benzyl). The peptides were cleaved from the Co-poly (styrene, 2% divinylbenzene) benzhydrylamine resin using hydrofluoric acid/10% anisole (v/v) at 4°C for 45 min. The solvent was removed under reduced pressure at 4°C. The resin was washed with ether and the peptide was extracted with trifluoroacetic acid (4x5 ml). The combined washes were evaporated and the residue redissolved in water and lyophilized. Cleavage of the peptides from the benzhydrylamine resin provides a carboxy-terminal amide.

The peptides were purified by reversed phase high performance liquid chromatography using a Synchropak RP-P C-18 analytical column (Synchrom Inc., Linden, Ind., 250x4.1 mm I.D.). Solvent A was 0.1% trifluoroacetic acid/water and solvent B was 0.05% trifluoroacetic acid/acetonitrile. A linear AB gradient of 1% B/min was used at a flow rate of 1 ml/min. The absorbance was recorded at 210 nm.

9. Amino acid analysis.

Amino acid analysis was performed on a Durrum D-500 amino acid analyser, after hydrolysis of the peptides in 6N HCl/0.1% phenol in sealed evacuated tubes at 110 degrees for 16-24 hours. The mean molar ratio of all accurately measurable amino acids was used to calculate the concentration of purified peptide.

C. RESULTS

Since F and ColB2 pilin monomeric subunits of pili have different amino termini (9,12) and are serologically distinct (17), the amount of cross-reactivity between polyclonal antisera raised in rabbits against F and ColB2 pili was assayed using a competition enzyme-linked immunosorbant assay (ELISA) (see Materials and Methods). This assay allows both a quantitative measure of antigen specific antibodies in polyclonal antisera as measured by the I_{max} (maximum inhibition), and a measure of the relative strength of the antigen-antibody interaction as determined by the inhibitor concentration required to give 50% inhibition (I_{50}). Serial dilutions of the competing antigen being tested are incubated with an appropriate amount of antiserum, after which the antigen-antiserum mixture is added to a microtiter plate precoated with pili (2 ug/ml). If the antigen in question reacts strongly with the antibodies in the antiserum, most antibodies are not available to bind to native pili in the microtiter wells, resulting in low ELISA readings (A_{405}). If there is little or no cross-reactivity, most of the anti-pilus antibodies will be free to react in the assay, and high ELISA readings (A_{405}) occur.

When anti-ColB2 pili antiserum was tested against F pili in a competition ELISA, an inhibition of about 20% was seen (see Fig. IV.1), as compared to ColB2 pili which was 100% . Similarly, when the reciprocal competition ELISA was done, it was found that the anti-F antiserum gave an I_{max} of 100% with the F pili, and an I_{max} value of 15% with heterologous ColB2 pili (Fig. IV.1).

Lipopolysaccharide (LPS) is a highly immunogenic component of bacterial cell walls. It was possible that some LPS from E.coli was contaminating the pili preparations and causing the cross reaction between the two pilus types. To prove that this cross reaction was not due to copurified lipopolysaccharide, antiserum was preabsorbed with E.coli JC6256 and then retested in a competition ELISA. No change in their cross reactivity was found. Also, purified LPS from the background strain E.coli JC6256 (kindly provided by E. Worobec, Univ. of Alberta, Edmonton) was tested in a competition ELISA and was found not to cross-react with polyclonal anti-pilus antiserum to any detectable extent (data not shown). This suggests that the cross reactivity between F and ColB2 pili is due solely to the pilus proteins.

Immunoblots of purified pilin separated on a 20% SDS polyacrylamide gel and electrophoretically transferred to nitrocellulose were tested with both antisera and found to verify the competition ELISA results (data not shown). It was found that bands of cross reacting pili were about 1/4 as intense as bands of homo-specific pili, suggesting approximately 15-25% antigenic cross reactivity between these two pilus proteins.

Since these two proteins are similar in sequence, yet antigenically distinct, the amino terminal region was examined to determine whether it may be responsible for the antigenic variation between these two pilus types. In another unrelated conjugative system, this laboratory has shown that the major antigenic determinant of pED208 pilin is located within the first five amino acids of the acetylated amino terminus (30). An analogous situation could prevail with F and ColB2 pilins.

F and ColB2 pilins contain a lysine residue at positions 17 and 15 respectively (Table IV.1). Tryptic cleavage of the pilin protein results in the release of the N-terminal peptide fragment and two other smaller fragments (9). The majority of the protein remains insoluble, and can be easily removed. To isolate the amino-terminal tryptic peptide of F or ColB2 pilin, tryptic digests of pilin (250 nmol) were fractionated on a Biogel P2 column (Biorad, Richmond, Ca.), using 0.05 N NH_4HCO_3 pH 8.0 as a buffer. The first peptide eluted after the void volume was pooled, concentrated by lyophilization and analyzed for purity and quantity by amino acid analysis assuming the peptide contained a single lysine residue. This peptide is referred to as F(1-17) or ColB2(1-15), depending on the type of pilin digested.

The antigenicity of the peptide was estimated using a competition ELISA. Typically 20-100 nmol/ml were used as the initial concentration of inhibiting peptide in the assay. The tryptic peptide ColB2(1-15) removed about 80% of the ColB2 pilus-specific antibodies (I_{max}=80%) (Fig. IV.2) and less than 25% of the F pilus-specific antibodies in competition ELISAs using homologous antisera (data not shown). No other fractions from the Biogel P2 column gave detectable activity in the ELISA. This suggested that the major antigenic determinant of the ColB2 pilin is probably localized within the first 15 amino acid residues of the protein.

This amino terminal peptide can be cleaved into two smaller peptides of 7 and 8 amino acid residues (ColB2(1-7) and ColB2(8-15)) by cyanogen bromide cleavage at the methionine at position 7 (see Materials and Methods). This results in the conversion of this methionine to homoserine. When the resulting mixture of the two peptides was tested in a competition ELISA, it was found to compete as well as the intact peptide ColB2(1-15), i.e., the I_{max} was 80% (Fig. IV.2). When the equivalent tryptic peptide from F pilin, F(1-17), which had been treated with cyanogen bromide to yield the F(1-9)/F(10-17) mixture was tested against ColB2 pilus-specific antiserum, an I_{max} value of less than 20% was obtained (Fig. IV.2). This suggests that the amino terminal peptide ColB2(1-7) contains the type-specific epitope, since the cyanogen bromide fragments F(10-17) and ColB2(8-15) are identical.

To confirm the foregoing observation, synthetic peptides were constructed to represent the amino terminal regions of F and ColB2 pilin. Peptides corresponding to the first 8 amino acids of F pilin F(1-8) and the first 6 amino acid residues of ColB2 pilin ColB2(1-6) were synthesized and purified according to the procedure in Materials and Methods. Both synthetic peptides were then tested in a competition ELISA using F and ColB2 pili and their homospecific antisera. The results are shown in Figure IV.3. It may be seen that the F(1-8) peptide was capable of removing approximately 80% of the F-specific antibodies in anti-F anti- serum, while the ColB2(1-6) peptide removed less than 20% of the anti-F pilus antibodies. In the case of ColB2-specific antiserum, approximately 60% of the antibodies were removed by ColB2(1-6) at a concentration of 250 nmol/ml, while inhibition by F(1-8) was negligible. It is of interest that the ColB2(1-6) peptide reacted with ColB2-specific antibodies over a very broad concentration range of peptide. However, the amount of inhibition was still increasing in a dose-related manner even at a peptide concentration of 250 nmol/ml, suggesting that this peptide may not contain the entire epitope.

The relative affinity of an antibody for an antigen can be expressed as the I_{50}, i.e., the amount of competing antigen required to cause 50% inhibition of the interaction of the pilus-specific antiserum with homologous pili in the microtiter wells of the competition ELISA (23). The I_{50} values obtained from Figs. 1-3 are summarized in Table 2. It may be seen that in the case of F-antiserum, the I_{50} for both F pili and the F(1-8) peptide was approximately 10 pmol/ml, indicating that the entire antigenic determinant for a subset of polyclonal antibodies specific for the amino terminal region of F pilin is contained in the peptide F(1-8). One or more other classes of antibodies recognise other separate epitopes which comprise the remaining 20% of the F pilus-specific antibodies.

The case of ColB2 pili and its homospecific antiserum is more complex. The I_{50} for whole pili was found to vary from 80-250 pmol/ml (the tendancy of pili to aggregate is reflected in the widely varying value for its I_{50}) and the inhibition curve reaches a maximum of 100% as expected. However, the I_{50} for the native peptide ColB2(1-15) was found to be 5000 pmol/ml. Surprisingly, a I_{50} value of only 1000 pmol/ml was obtained after cleavage of the peptide with cyanogen bromide. Both peptides were capable of giving a maximum inhibition value of 80%, indicating that they are both reactive with the class of antibodies specific for the amino terminus of ColB2 pilin and that the entire antigenic determinant is contained within these peptides. The variation in the I_{50} for these peptides is probably due to a conformational component of the antigenic determinant that is destroyed by proteolytic and chemical cleavage. It is interesting that the antigen-antibody interaction appears to be stronger in the case of the peptide mixture ColB2(1-7)(8-15) obtained by CNBr cleavage than the uncleaved peptide ColB2(1-15) (a 5-fold decrease in the amount of CNBr- treated peptide mixture required to give 50% inhibition in the competition assay). This can be explained by the shorter peptides being able to adopt more favorable conformations for interaction with the antibody molecule.

As seen in Fig. IV.3, the inhibition curve for the synthetic peptide ColB2(1-6) was extremely shallow, yielding an I_{50} of 100,000 pmol/ml. The maximum inhibition plateau was not reached within the limits of the experimental design. This would seem to indicate that the ColB2(1-6) peptide is an incomplete sequential determinant and that the residues 7-15 contribute to this epitope.

D. DISCUSSION

The major antigenic determinant in three types of conjugative pili has been localized in the amino terminal region of the pilin protein (Table IV.2). In the pED208 system, the acetyl group and the first five amino acids are involved in antigen-antibody recognition (30). Using synthetic peptide analogs, we have reported that the important contact points in antibody binding of this epitope involve the side chains of Leu3 and Leu4, the hydroxyl group of Thr1 and the methyl group of the N-acetyl moiety (30).

The sequences of F and ColB2 amino termini are completely different from that of pED208 except for the acetyl group at the amino terminus. As reported in this paper, there is little (<20%) cross-reactivity between F(1-8) and ColB2(1-6) synthetic peptides, even though they share four amino acids (-gly-gln-asp-leu) at their carboxy termini. This is consistent with the antigenic determinant being located at the amino terminus of the peptide, as is the case with pED208. It is also noteworthy that the shared N-acetyl-alanine is insufficient to promote strong cross-reactivity as detected by competition ELISA. This is similar to the pED208 system, where it was found that although the acetyl methyl and threonine hydroxyl groups of the N-terminal residue contribute significantly to the EDP208 antigen-antibody interaction, N-acetylthreonine alone was unable to cause detectable inhibition in the competition ELISA (30).

Ippen-Ihler et al.(14) have reported that the 13,000 dalton precursor form of F pilin, propilin, is only weakly reactive in immunoblots with anti-F pilus antiserum, whereas mature pilin is strongly reactive. The weak reactivity of propilin may be attributable to the secondary antigenic determinant which reacts with 20% of the antiserum, or it may also reflect a partial reactivity of the antiserum with an uncleaved, unacetylated N-terminal epitope.

In 1970, Lawn and Meynell (19) identified four serotypes for F-like pili: F, R538-1, R1-19, and R100-1. They found that F, R538-1, and R100-1 were weakly cross-reactive, while R538-1 and R1-19 were strongly cross-reactive with each other. Since F and R100-1 did not cross-react with R1-19, their serological evidence suggested that F, R538-1 and R100-1 pili each contain a unique major antigenic determinant and R1-19 lacks a second determinant common to F, R538-1 and R100-1. The results presented in this paper, and a previous communication (30), show that the type-specific determinants observed for F, ColB2 and EDP208 pili are located at the N-terminus of the pilin protein. Additional studies in this laboratory (to be reported elsewhere) have revealed that a unique epitope also exists at the N-terminus of R100-1 pilin, while the N-terminus of R1-19 and R538-1 pilins are identical to that of ColB2. These observations are consistent with the findings of Lawn and Meynell (19), and provide a molecular basis for the serological specificity they observed.

In the pilus systems studied to date, there appears to be one area in each pilin protein that is highly immunogenic. In conjugative pili the major determinant is located at the N-terminus. In pili which serve as adhesins for mammalian cells, e.g., gonococcal, pseudomonas, and K88, the immunodominant antigenic determinants have been shown to be nearer the center of the polypeptide (18,25,26). There thus seems to be a pattern for pilus antigenicity emerging, although additional pilus types will have to be characterized before any generalizations can be made.

TABLE IV.1

Amino terminal tryptic peptide sequences of EDP208, F and ColB2 Pilins.

Pilin Type	Amino terminus sequence	Reference
EDP208	N-Ac-thr-asp-leu-leu-ala-gly-gly-lys-asp-val-asp-lys	(11)
F	N-Ac-ala-gly-ser-ser-gly-gln-asp-leu-met-ala-ser-gly-asn-thr-thr-val-lys	(12)
ColB2	N-Ac-ala-gln-gly-gln-asp-leu-met-ala-ser-gly-asn-thr-thr-val-lys	(9)

TABLE IV.2

Competition ELISA Values (I_{max} and I_{50}) Obtained with Synthetic and Tryptic Amino Terminal Peptides of F and ColB2 Pilin.

	F-Specific Antiserum		ColB2-Specific Antiserum	
Competitor Antigen	I_{max}[a] (Percent)	I_{50}[b] (pmol/ml)	I_{max}[a] (Percent)	I_{50}[b] (pmol/ml)
Homologous pili	100	10	100	80-250
ColB2(1-15)	-	-	80	5,000
ColB2(1-15)/CNBr[c]	-	-	80	1,000
ColB2(1-6)	15	500[e]	>60	100,000
F(1-17)/CNBr[d]	-	-		10,000[e]
F(1-8)	80	10	-	-

[a] Maximum percent inhibition achieved with a given competitor antigen.

[b] Concentration of competitor in pmol/ml required to achieve a 50% inhibition.

[c] Peptide cleaved with CNBr to yield a mixture of ColB2(1-7)(8-15). Met7 was converted to homoserine during the CNBr cleavage.

[d] Peptide cleaved with CNBr to yield a mixture of F(1-9)(10-17). Met9 was converted to homoserine during the cleavage.

[e] Concentration of competitor in pmol/ml required to achieve a maximum inhibition of 15%.

FIGURE IV.1

Competition ELISAs of purified F and ColB2 pili assayed with rabbit anti-pilus polyclonal sera directed against ColB2 pili (A) and F pili (B). The competitor initial concentrations were 20 nmol/ml for F and ColB2 pili: two-fold dilutions were made to construct the inhibition curve.

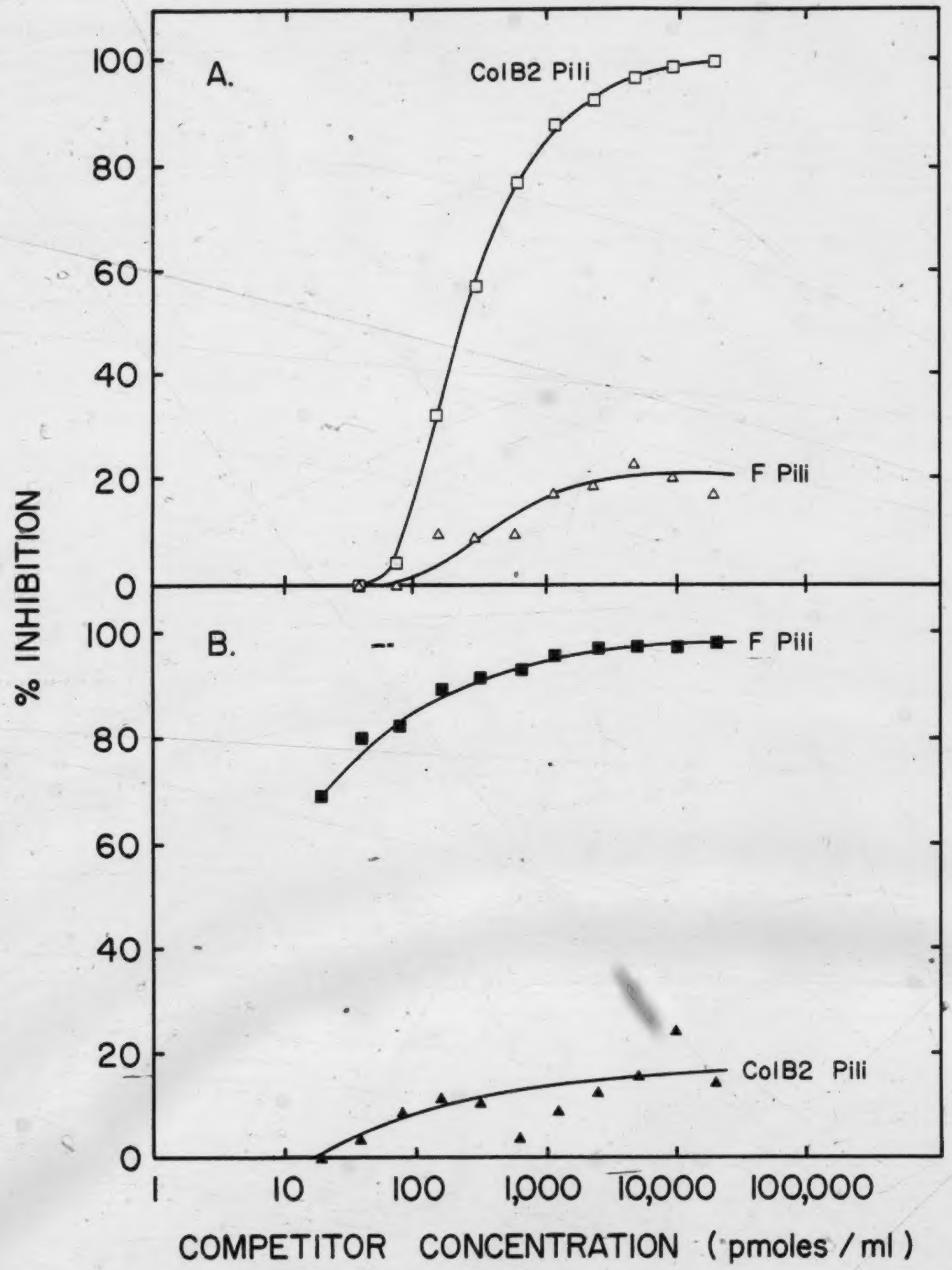

A.
ColB2 Pili
F Pili
B.
F Pili
ColB2 Pili
% INHIBITION
100
80
60
40
20
0
1
10
100
1,000
10,000
100,000
COMPETITOR CONCENTRATION (pmoles / ml)

FIGURE IV.2

Competition ELISAs of amino terminal tryptic peptides treated with cyanogen bromide (CNBr). The peptide mixtures were assayed with anti ColB2 pilus antiserum. The ELISA was carried out as described in Materials and methods using initial competitor concentrations of 100,000 pmol/ml F(1-17)/CNBr [mixture of F(1-9)(10-17)]; 200,000 pmol/ml ColB2(1-15) and 13,000 pmol/ml ColB2(1-15)/CNBr [mixture of ColB2(1-7)(8-15)]. Two-fold dilutions were used to construct the inhibition curve. CNBr treated peptide mixtures contained homoserine in place of methionine (see text for details).

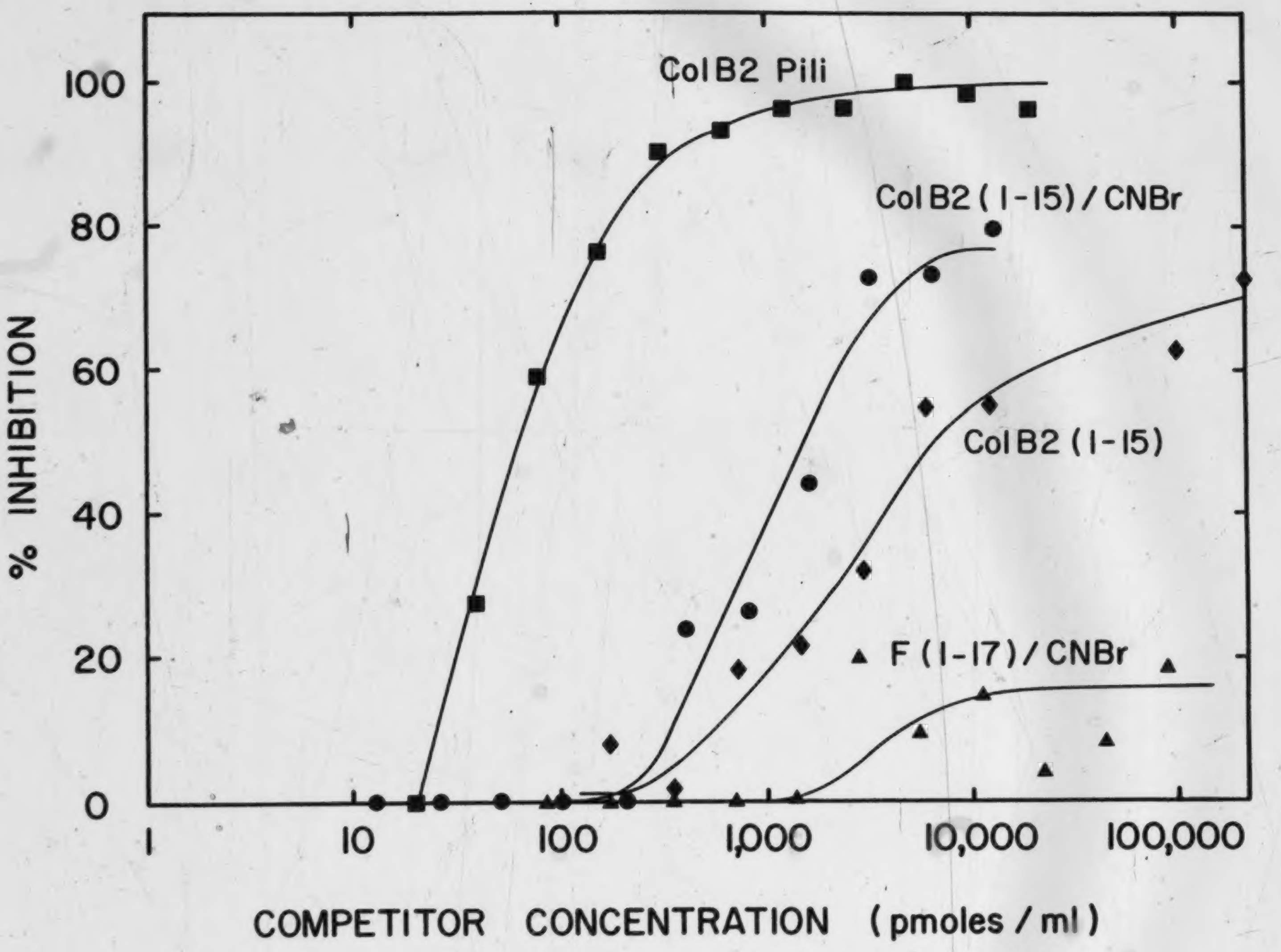
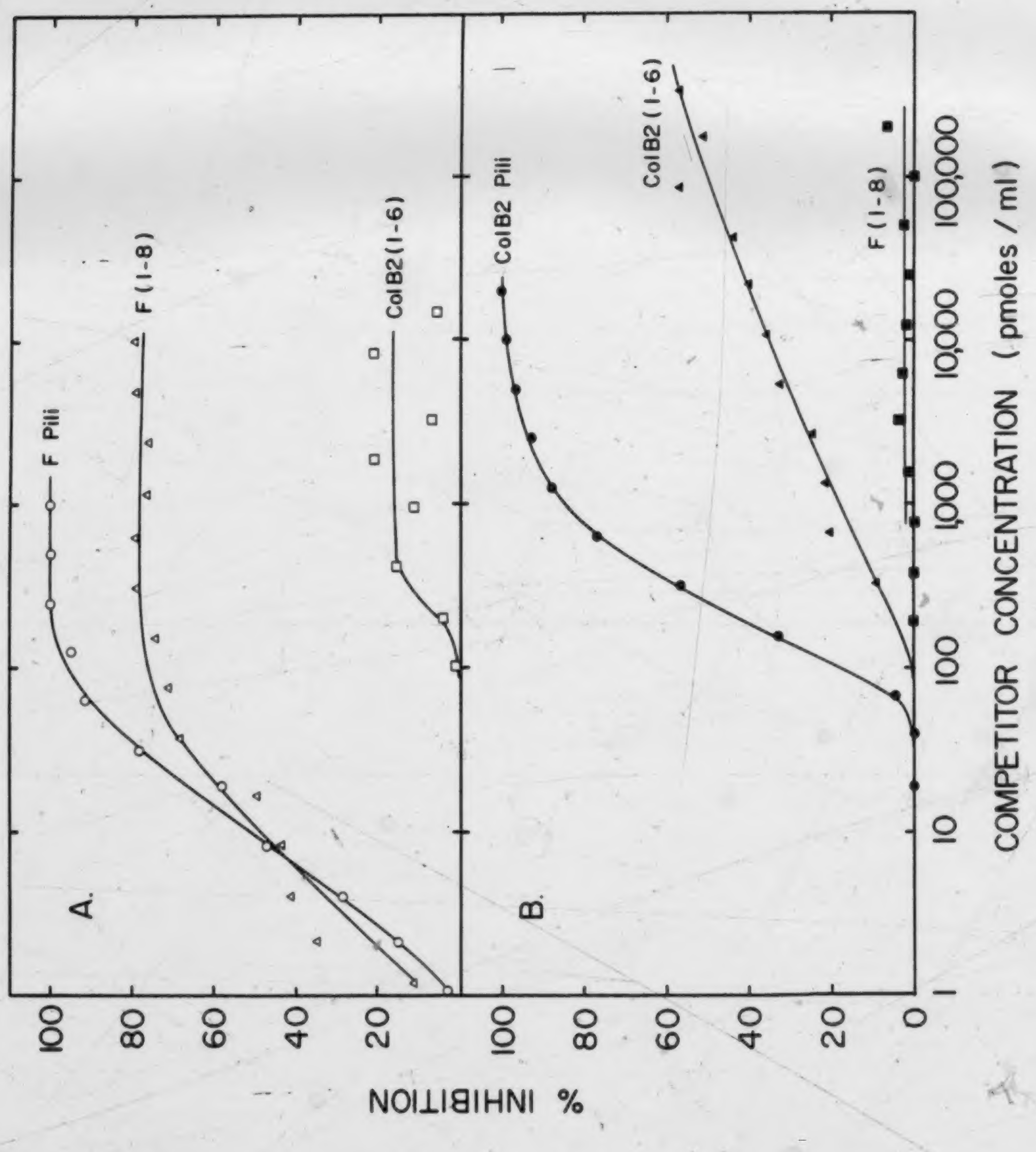

% INHIBITION
100
80
60
40
20
0
1
10
100
1,000
10,000
100,000
COMPETITOR CONCENTRATION (pmoles / ml)
Col B2 Pili
Col B2 (1-15) / CNBr
Col B2 (1-15)
F (1-17) / CNBr

FIGURE IV.3

Competition ELISAs of amino terminal synthetic peptides tested against F (A) and ColB2 (B) pilus specific polyclonal antisera. The initial compeptitor concentrations were: F pili 1000 pmol/ml; ColB2 pili 20,000 pmol/ml; F(1-8) 10,000 pmol/ml; ColB2(1-6) 342,000 pmol/ml. Two-fold dilutions were made to construct the inhibition curve.


A.
B.
F Pili
F (1-8)
ColB2 (1-6)
ColB2 Pili
ColB2 (1-6)
F (1-8)
% INHIBITION
100
80
60
40
20
0
1
10
100
1,000
10,000
100,000
COMPETITOR CONCENTRATION (pmoles / ml)

E. BIBLIOGRAPHY

1. **Achtman, M., and R. Skurray**. 1977. A redefinition of the mating phenomenon in bacteria. p. 233-279. In J.L. Reissig (ed.), Microbial Interactions, Receptors and Recognition. series B, vol. 3. Chapman & Hall, Ltd. London.

2. **Armstrong, G.D., L.S. Frost, H.J. Vogel, and W. Paranchych**. 1981. Nature of the carbohydrate and phospholipid associated with ColB2 and EDP208 pilin. J. Bacteriol. **145**:1167-1176.

3. **Bradley, D. E. and E. Meynell**. 1978. Serological characteristics of pili determined by the plasmids R711b and Flac. J. Gen. Microbiol. **108**:141-149.

4. **Brinton, C.C**. 1971. The properties of sex pili, the viral nature of conjugal genetic transfer systems, and some possible approaches to the control of drug resistance. Critical Rev. Microbiol. 1:105-160.

5. **Bukhari, A., J. Shapiro and S.L. Adhya**. 1977. DNA Insertion Elements, Plasmids and Episomes, p. 601-704. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

6. **Caro, L.G. and M. Schnoss**. 1966. The attachment of male-specific bacteriophage f1 to sensitive strains of Escherichia coli. Proc. Nat. Acad. Sci. U.S.A. **56**:126-132.

7. **Clowes, R.C**. 1972. Molecular structure of bacterial plasmids. Bacteriol. Rev. **36**:361-405.

8. **Crawford, E.M. and R.F. Gesteland**. 1964. The adsorption of bacteriophage R17. Virology. **22**:165-167.

9. **Finlay, B. B., L. S. Frost, and W. Paranchych**. 1984. Localization, cloning, and sequence determination of the conjugative plasmid ColB2 pilin gene. J. Bacteriol. **160**:402-407.

10. **Folkhard, W., K. R. Leonard, S. Malmsey, D. A. Marvin, J. Dubochet, A. Engel, M. Achtman, and R. Helmuth**. 1979. X-ray diffraction and electron microscope studies on the structure of bacterial F pili. J. Mol. Biol. **130**:145-160.

11. **Frost, L.S., G.D. Armstrong, B.B. Finlay, B.F.P. Edwards, and W. Paranchych**. 1983. N-terminal amino acid sequencing of EDP208 conjugative pili. J. Bacteriol. **153**:950-954.

12. **Frost, L. S., W. Paranchych, and N. S. Willetts**. 1984. DNA sequence of the F traALE region that includes the gene for F pilin. J. Bacteriol. **160**:395-401.

13. **Hausmann, C. and R.C. Clowes**. 1971. ColB2-K77, a fertility-repressed F-like sex factor. J. Bacteriol. **107**:900-906.

14. **Ippen-Ihler, K., D. Moore, S. Laine, D.A. Johnson and N.S. Willetts**. 1984. Synthesis of the F-pilin polypeptide in the absence of F traJ product. Plasmid. **11**:116-129.

15. **Kasper, C.B**. 1975. Fragmentation of proteins for sequence studies and separation of peptide mixtures. Mol. Biol. Biochem. Biophys. **8**:114-118.

16. **Kennedy, N., L. Beutin, M. Achtman, R. Skurray, W. Rahmsdorf and P. Herrlich**. 1977. Conjugation proteins encoded by the F sex factor. Nature (London). **270**:580-585.

17. **Ketyi, I. and I. Orskov**. 1969. Studies on the antigen structure of sex fimbriae carried by a strain of Shigella flexnerii 4b. Acta. Pathol. Microbiol. Scand. **77**:249-308.

18. **Klemm, P**. 1981. The complete amino acid sequence of the K88 antigen, a fimbrial protein from Escherichia coli. Eur. J. Biochem. **117**:617-627.

19. **Lawn, A.M. and E.E. Meynell**. 1970. Serotypes of sex pili. J. Hygiene. **68**:683-694.

20. **Minkley, E.G., S. Polen, C.C. Brinton and K. Ippen-Ihler**. 1976. Identification of the structural gene for F pilin. J. Mol.Biol. **108**:111-121.

21. **Moore, D., B.A. Sowa and K. Ippen-Ihler**. 1981. The effect of tra mutations on the synthesis of the F-pilin membrane polypeptide. Mol. Gen. Genet. **184**:260-264.

22. **Moore, D., B.A. Sowa, and K. Ippen-Ihler**. 1982. A new activity in the Ftra operon which is required for F-pilin synthesis. Mol. Gen. Genet. **188**:459-464.

23. **Nieto, A., A. Gaya, M. Jansa, C. Moreno and J. Vives**. 1984. Direct measurement of antibody affinity distribution by hapten-inhibition enzyme immunoassay. Mol. Immunol. **21**:537-543.

24. **Paranchych, W.** 1975. Attachment, ejection and penetration stages of the RNA phage infectious process. p.85-111. *In* N. Zinder (ed.), RNA Phages. Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.

25. **Schoolnik. G.K., J.Y. Tai, and E.C. Gotschlich**. 1982. Receptor binding and antigenic domains of gonococcal pili, p. 312-316. *In* D. Schlessinger (ed.), Microbiology-1982. American Society for Microbiology, Washington, D. C.

26. **Watts, T. H., P. A. Sastry, R. S. Hodges, and W. Paranchych**. 1983. Mapping of the antigenic determinants of *Pseudomonas aeruginosa* PAK polar pili. Infect. Immun. **42**:113-121.

27. **Willetts, N. S., and R. Skurray**. 1980. The conjugative system of F-like plasmids. Annu. Rev. Genet. **14**:41-76.

28. **Willetts, N.S. and B. Wilkins**. 1984. Processing of plasmid DNA during bacterial conjugation. Microbiol. Rev. **48**:24-41.

29. **Worobec, E. A., A. K. Taneja, R. S. Hodges, and W. Paranchych**. 1983. Localization of the major antigenic determinant of EDP208 pili at the N-terminus of the pilus protein. J. Bacteriol. **153**:955-961.

30. **Worobec, E. A., W. Paranchych, J. M. R. Parker, A. K. Taneja, and R. S. Hodges**. 1985. Antigen-antibody interaction: the immunodominant region of EDP208 pili. J. Biol. Chem. **260:938-943**

Chapter V

NUCLEOTIDE SEQUENCES OF THE R1-19 PLASMID TRANSFER GENES *traM*, *finP*, *traJ*, and *traY* AND *traYZ* PROMOTER[1]

A. INTRODUCTION

Resistance R factors are large, self-transmissible plasmids which encode antibiotic resistance genes. R1 is a member of the IncFII plasmid incompatibility group and contains a 35 kilobase (kb) transfer (*tra*) region which is highly homologous to that in plasmid F and other IncF plasmids (28). This *tra* region encodes the gene products necessary for pilus production, surface exclusion, and conjugal DNA metabolism and is controlled by several regulatory elements. Of the *tra* regions in the IncF plasmids, the F plasmid *tra* region has been studied most extensively (34,35). R1*drd*-19 (R1-19) is a derepressed mutant of R1 (*finO*) and expresses its conjugation system constitutively (20).

Although the *tra* regions of the F and R1-19 plasmids are homologous, they are not identical. Sharp *et al*. (28)

[1] 1. A version of this chapter has been accepted for publication. Finlay, B.B., L.S. Frost, and W. Paranchych. 1986. J. Bacteriol. May Issue

observed two small regions of nonhomology which mapped near the beginning of the tra regions of these two plasmids (6). Ostermann et al. (23) demonstrated that R1-19 and F plasmids have both homologous and nonhomologous sequences within their origin of transfer (oriT) regions. Frost et al. (10) found that, although the traALE regions of plasmids R1-19 and F were very similar, the DNA sequences diverged immediately upstream of traA (the pilin gene). The present DNA sequencing studies of the region between the oriT and traA genes from R1-19 confirmed that these regions are nonhomologous in the R1-19 and F plasmids.

This region contains several tra genes (traM, finP, traJ, traY) and the traYZ promoter. The traM product is essential for DNA transfer and may be involved in the triggering of DNA transfer and conjugative DNA synthesis by binding to a plasmid specific region of the oriT gene (36). Four alleles of the traM gene have been described for IncF plasmids (F-like, R1-like, R100-like, and ColB4-like). Similarily, these traM variants are found with plasmid specific oriT sequences and four traY alleles (34). The traY product is thought to be a component of the traYZ endonuclease, which may be responsible for strand-nicking at the oriT gene (36). It has been suggested that the traY product may be involved in plasmid-specific oriT recognition whereas the traZ protein is probably responsible for nicking the DNA (34). The F plasmid *traJ* product is required for expression of the *traYZ* and *traM* operons of plasmid F (12,21). It has been shown that the F plasmid *traJ* product increases transcriptional levels at the *traYZ* promoter (21), which is located immediately downstream of the *traJ* gene (9). The *traYZ* transcript proceeds from *traY* to *traZ* and is approximately 32 Kb long. Four *traJ* alleles in the IncF plasmid group have been identified, three of which correspond to the four *traMY*/*oriT* alleles. [Type II (ColB4-like) and type III (R1-like) *oriT* alleles both have a type III *traJ* allele and R-386 has a unique *traJ* allele (34).] Expression of the *traJ* gene is repressed by the *finOP* (fertility inhibition) system. The *finO* gene is not plasmid specific, and there are at least six different *finP* types in the IncF group (34). The *finP* gene is located in the region between the *traM* and *traJ* genes, being transcribed in the opposite direction from that of *traJ*, while the *finO* gene is not linked with the *tra* region. The lack of either the *finO* or *finP* gene relieves *traJ* repression and allows concomitant *traYZ* operon expression, producing a derepressed state.

In this chapter the nucleotide sequences of the R1-19 *traM*, *finP*, *traJ*, and *traY* genes and the *traYZ* promoter are presented, and compared to the related, but not identical genes from the F conjugative plasmid.

B. MATERIALS AND METHODS

1. Bacterial strains and plasmids.

The R1*drd*-19 plasmid was carried in the host strain *Escherichia coli* JC 6256 (F^- *lac trp*). pUC8 recombinants were transformed into *E. coli* JM83 (33) and galactokinase (*galK*) chimeras containing the promoter assessment vector pKO4 were transformed into *E. coli* ED 5363 (*gal*) (19). pBF402 is an 8 kb *Eco*RI fragment from R1-19 ligated into pUC8 and is equivalent to the *Eco*RI f6 fragment from the F plasmid. pRS27 is the *Eco*RI f6 fragment from plasmid F ligated into pSC101 (30). pBF403 is a 320 basepair R1-19 *Sau*3A fragment (Fig. V.2, nucleotides 1602-1922) ligated into pKO4/*Bam*HI (19) and pBF302 consists of the 682 basepair *Sau*3A fragment from plasmid F (containing the *traYZ* promoter)(9) and the vector pKO4. pBF405 is a recombinant containing the 1847 basepair R1-19 *Aha*III fragment (Fig. V.2, nucleotides 200-2047) and pACYC184 (4).

2. Plasmid purification.

R1-19 plasmid and the chimeras were isolated and purified by the method of Humphreys *et al*. (15) and small amounts of chimeric plasmids were obtained by the method of Birnboim and Doly (3).

3. Recombinant DNA techniques and DNA sequencing.

Recombinant DNA techniques and DNA sequencing were performed as described previously (8,10), except that M13mp18 and M13mp19 were used as the M13 vectors (37).

4. Galactokinase (galK) assay.

The galK assay was done exactly as described by McKenney *et al*. (19), except that the specific activity of the undiluted [^{14}]C-galactose was 59.6 mCi/mmol. Units were calculated as described previously (19), and are expressed as nanomoles galactose phosphorylated/min/OD_{650}.

C. RESULTS

1. Cloning and Sequencing Strategies.

When R1*drd*-19 plasmid DNA/*Eco*RI was separated on an agarose gel, immobilized on nitrocellulose, and probed with a 1.4 kb *Pst*I fragment containing the *traALE* genes of the F plasmid transfer region (11), a single R1-19 *Eco*RI fragment (8 kb) showed homology (data not shown). This fragment corresponded to the 8 kb f6 *Eco*RI fragment from F plasmid which contains the F plasmid *traMJYALEKBP*, *finP* and *oriT* genes (35). This *EcoRI* fragment from R1*drd*-19 was cloned into the vector pUC8 (33) and called pBF402. The 1.26 kb *Bgl*II/*Pst*I fragment and the adjacent 2.1 kb *PstI* fragment were subcloned into pUC8 and sequenced by using the M13 dideoxy sequencing method (26). A restriction map, the sequencing strategy and the putative gene products found in the *BglII*/*PstI* fragment and in the first half of the 2.1 kb *PstI* fragment are shown in Fig. V.1. The remainder of the *PstI* fragment contains the *traALE* genes and the nucleotide sequence of the *traA* gene was reported previously (10). The R1-19 *traL* gene and most of the *traE* gene have been sequenced and are nearly identical to the corresponding sequence from the F plasmid (B. Finlay, unpublished results). Fig. V.2 contains the nucleotide sequences and the predicted products of the *traM*, *finP*, *traJ* and *traY* genes. (Included in this figure is the 450 nucleotide sequence published previously which contains the R1 *oriT* region (23). This sequence corresponds to nucleotides 25-474 in Fig. V.2.). Recently Koronakis *et al*. (17) determined the nucleotide sequence of the *traM* gene from the R1 plasmid. This sequence is identical to the R1-19 *traM* sequence presented here except for the adenine at position 893, which is not present in the R1 sequence. This difference would not alter the predicted gene products.

By using heteroduplex mapping, Sharp *et al*. (28) found that the *tra* operons of the F and R1 plasmids are highly homologous whereas the remainder of the plasmids show little similarity. Sequencing studies on the R1-19 pilin gene have demonstrated that although the traALE sequences from the F and R1-19 plasmids are very similar, these sequences diverge immediately upstream of traA (10). The heteroduplex mapping results of Sharp *et al*. reveal a 1.2 kb region of nonhomologous DNA upstream of the predicted location of the traA gene [64 kb (28); 69 kb (35)]. The nonhomologous region corresponds to map positions 60.0 to 61.2 in the R1 plasmid and 62.7 to 63.9 in the F plasmid (28). Upstream of this nonhomologous sequence is 0.4 kb of homologous DNA (map positions 60.0 to 59.6 in R1 and 62.7 to 62.3 in F, again using the map units of Sharp *et al*.), preceded by 0.4 kb of nonhomologous DNA (map positions 59.6 to 59.2 in R1 and 62.3 to 61.9 in F). There are no other nonhomologous areas upstream until well beyond the oriT gene, the beginning of the tra region.

Comparing the sequence from the R1-19 plasmid (Fig. V.2) with the corresponding sequence from the transfer region of the F plasmid (9,31,32) revealed strong homology from nucleotides 1 to 173 in R1-19 which includes the three major nick sites for the oriT region (32) (located after positions 124, 127, and 130 in Fig. V.2). From nucleotides 174 to 592 there was little similarity, but from 593 to 1053 there was again strong homology, except for the area between 854 and 936 where the homology was limited. From nucleotides 1054 to 2091 there was little homology, although the regions encoding the carboxyl termini of the traY gene products were similar. These results correlated well with the heteroduplex studies of Sharp et al (28), accounting for the two nonhomologous regions seen by these investigators.

2. R1-19 traM gene.

The R1-19 traM gene encodes a 127 amino acid protein (14,448 daltons) which shows considerable homology with the traM gene product from plasmid F [(31); Fig. V.3]. The DNA sequences encoding the amino terminal regions of the traM gene products from the F and R1-19 plasmids differed quite significantly and this difference was reflected in the protein sequences. Beyond residue 40 (nucleotide 592) there was a high level of protein homology, although there was a charge difference (Lys/Glu) at residue 123.

The traM gene product from plasmid F is an inner membrane protein which does not undergo proteolytic processing (1). Although the amino terminal protein sequences of the traM gene products from plasmids R1-19 and F are different, neither resembles the classical hydrophobic signal sequence (16), and it is not known how this protein enters the inner membrane.

Six nucleotides upstream of the traM start codon is a strong (five nucleotide) ribosome binding site (29). The RNA transcript and putative promoter for the F traM gene have been mapped (31), and there was no homologous sequence in the R1-19 plasmid. There was a possible promoter upstream of the traM gene in R1-19 (Fig. V.2) which resembled the consensus promoter sequence (13,25).

In plasmid F, immediately downstream of the traM stop codon, there is a region of dyad symmetry, followed by several thymine residues, which may be responsible for transcriptional termination of the traM transcript (31). The R1-19 plasmid had a small (5 base pair) region of dyad symmetry starting at position 869 (Fig. V.2). It is possible that this was the R1-19 traM transcription terminator, but the resemblance to a rho-independent consensus terminator sequence was weak (25).

3. R1-19 finP gene.

The regions encoding the finP genes in R1-19 and F plasmids were very similar. The inverted repeat sequences thought to be involved in finOP binding [fisO; (21)] were conserved, but the 7 nucleotides between these areas of dyad symmetry (positions 1000 to 1006) were different. These changes would affect the proposed 24 amino acid finP peptide (21) (starting at position 1027 and proceeding in the direction opposite to traJ transcription to position 955) by changing Thr-6 Ser-7 to Arg-6 Pro-7. If the finP product is an RNA molecule, these changes would affect the loop in a stem and loop structure, possibly providing the required finP plasmid specificity (see Discussion).

The proposed finP promoter in plasmid F is located within the sequence encoding the amino terminal area of the traJ protein, with transcription proceeding in the opposite direction to that of traJ transcription (21). Because DNA homology between the F and R1-19 plasmids disappeared at position 1054, this should have affected the predicted finP promoter. However, the putative Pribnow box [TACGAT (Fig. V.2)] was identical to that of plasmid F (31) and the -35 region (TTGACT) differed by only one nucleotide (T_{1086} in R1-19 is a G in plasmid F). The sequence between these areas showed no homology, nor did the surrounding regions. The opposite strand of the small inverted repeat previously mentioned as a possible termination signal for traM transcription (positions 886 to 869) may also serve as a finP terminator. When transcribed in the finP direction, this area of dyad symmetry was followed by several thymine residues, which are usually present in rho-independent terminators (25).

4. R1-19 traJ gene.

The F plasmid traJ gene product is required for transcription of the traYZ operon (34). The region encoding the traJ gene in plasmid F and the equivalent area in R1-19 showed no homology, which was in accordance with the findings of Alfaro and Willetts (2) who demonstrated that R1-19 was unable to complement an F plasmid traJ mutant (F*lac* *traJ90*). R1-19 encodes a 201 amino acid (23,241 dalton) polypeptide (positions 1050 to 1652) in this nonhomologous area.

In plasmid F, the gene product of *traJ* is required for *traYZ* operon transcription (35). Fowler *et al*. (9) demonstrated that a recombinant containing the *traJ* gene could stimulate the *traYZ* promoter approximately 2.1 fold above uninduced levels. To test whether the R1-19 plasmid has analogous *traJ*-like activity, a 1847 base pair R1-19 *Aha*III fragment (pBF405; positions 200 to 2047) which contained the *traM* and *finP* genes, the *traYZ* promoter, most of the *traY* genes and the putative *traJ* polypeptide was cloned into pACYC184. This chimera was able to increase transcription of the R1-19 *traYZ* (P_{YZ}) promoter 3.7 fold (see Table V.1). Based on this increased transcription and the fact that functions have been assigned to the remaining genes on this fragment, we suggest that this chimeric protein is the *traJ* equivalent in the R1-19 plasmid. It was similar in size to the F plasmid *traJ* protein [27,031 daltons; (9)]. The R1-19 *traJ* gene had ribosome binding and translational start sites which were homologous to those of the F plasmid *traJ* gene (GGAGGTTCCTATG). Except for the *finP* promoter, there was no homology at the DNA or protein level beyond position 1054 (Met-1) . F and R1-19 *traJ* proteins showed no similarity in the secondary structures predicted by the algorithms of Chou and Fasman

(5) or Hopp and Woods (14) (data not shown).

The F plasmid traJ protein is believed to be incorporated into the outer membrane of bacteria without proteolytic processing (1). Examination of the F plasmid protein sequence revealed neither a signal at the amino terminus, nor any apparent hydrophobic regions that would anchor the protein in the outer membrane (9). The R1-19 traJ sequence also showed no apparent signal sequence or hydrophobic domain. Fowler et al (9) suggested that the tertiary structure of the F plasmid traJ protein could bring together small hydrophobic regions which could then form a hydrophobic membrane-anchoring domain. The role of the traJ protein in the outer membrane is currently unknown, and the paradox of how an outer membrane protein can regulate a promoter remains unresolved.

5. The traYZ Promoter (P_{YZ}).

In the F plasmid the P_{YZ} promoter is located between the traJ and traY genes (9,21). Although there was no DNA homology between the R1-19 and F plasmids in this area, the gene products downstream of the traY gene were functionally interchangeable, and an analogous control system for R1-19 was expected. To test for a P_{YZ}-like promoter activity, a 320 bp Sau3A fragment (1602-1922), which extends from the end of the traJ gene to the beginning of the traY gene, was cloned into the galK expression vector pKO4/*Bam*HI (19) and assayed for promoter strength. This region had promoter activity by itself (Table V.1). However, when the R1-19 plasmid was present (i.e., a functional *traJ* gene is supplied) this promoter was stimulated threefold. When a chimera containing the *traJ* gene (pBF405) is present, transcription from this promoter was increased 3.7-fold over uninduced levels. These results are very similar to the results obtained with the F plasmid P_{YZ} promoter (9), which showed a 2.1-fold increase when the F plasmid *traJ* gene was present. (In our assay, a 1.9-fold increase in P_{YZ} transcription was observed.) This suggests that the DNA between the *traJ* and *traY* genes contains the R1-19 P_{YZ} promoter. The F plasmid *traJ* gene was unable to stimulate this promoter, and the R1-19 *traJ* gene was unable to affect the F plasmid P_{YZ} activity (see Table V.1), confirming that R1-19 and F plasmids have different *traJ*/P_{YZ} alleles (34).

The F plasmid P_{YZ} promoter has been sequenced (9) and encompasses a *Bst*EII site between the *traJ* and *traY* genes (21). The proposed -35 and -10 regions (CTGCAA and GAAGAT) do not resemble the procaryotic promoter consensus sequences (TTGACA and TATAAT) (13). This is not surprising because the P_{YZ} promoter must contain a site for *traJ* product interaction and positive control promoters usually differ significantly from the procaryotic promoter consensus sequence (24). When the R1-19 sequence between the *traJ* and *traY* gene was compared with the consensus promoter sequence, no homologous regions were found. There was also no DNA homology between the R1-19 and F plasmids in the P_{YZ} region. Therefore, the R1-19 P_{YZ} sequence remains undefined, but was localized between the *traJ* and *traY* genes on a 320 bp *Sau*3A fragment.

In the F plasmid, *traJ* product is not only required for *traYZ* operon transcription, but also possibly for *traM* transcription (12). *In vitro* studies on the *traM* promoter (P_M), which are contradictory to the above results, suggest that the *traM* promoter is constitutive (21). If this promoter was under *traJ* control, a homologous sequence between P_M and P_{YZ} that the *traJ* product could recognize would be expected. In the R1-19 plasmid , there is an identical 9 nucleotide sequence 131 nucleotides upstream of the *traM* gene (nucleotides 342 to 350) and 122 nucleotides upstream of the *traY* gene (nucleotides 1710 to 1718). These regions may be involved in *traJ* protein recognition and interaction with these two promoters.

6. R1-19 *traY* gene.

The F plasmid *traY* gene product, believed to be a component of the *traYZ* endonuclease (36), is encoded by the region of DNA between P_{YZ} and the *traA* gene (9). When the analogous area of the R1-19 plasmid was examined, a single open reading frame was found which spanned nucleotides 1832 to 2257 (75 residues). Although this region contained no AUGs, a GUG at nucleotide 1832 was preceded by a strong Shine-Dalgarno sequence [GAGGT; (29)] 7 residues upstream. Although AUG is the usual start codon, there are several examples of GUG being used as an initiation codon (18). This traY polypeptide was be 75 residues in length (9037 daltons) and could be aligned with F plasmid traY gene product (Fig. V.4) suggesting that it is the R1-19 traY gene product.

The product of the traY gene from the R1-19 plasmid was 44 amino acids shorter than that of the F plasmid leading to a truncated amino terminal region in the R1-19 protein. In addition, the amino termini of the two gene products were completely different. The protein homology between the traY products from plasmids F and R1-19, starting at residue 62 in the F traY protein, was reflected in the DNA sequence; there was extensive DNA homology in the latter half of the traY genes of plasmids F and R1-19 which continues into the traALE sequences (10).

D. DISCUSSION

In the IncF plasmid group, there are four oriT types [F-like, R1-like, ColB4-like and R100-like; (34)]. The sequences of F and R1-19 oriT genes have been published (23,32). There is a highly homologous region that contains the three proposed nick sites (32) corresponding to nucleotides 1 to 173 in Fig. V.2 followed by a sequence between nucleotide 174 and the traM gene that is nonhomologous and may be involved in binding of the plasmid-specific gene products of traY and traM as part of the process of DNA metabolism during DNA transfer (36). The traY gene products of the F, R1-19 and R100 plasmids are not interchangable (34), and because the traZ products of the R1-19 and F plasmids can complement each other (7), Willetts and Maule have suggested that the conserved traZ protein is the component of the traYZ endonuclease that actually nicks at the homologous oriT sequences whereas the plasmid specific traY product binds to the DNA at the nonhomologous sequence within the oriT region.

The traM gene product is also oriT specific and not interchangable among F-like plasmids (34). It is believed to be involved in triggering conjugal DNA replication, possibly by binding at or near the oriT region (36). Since the traM gene product is plasmid specific, it may also recognize this area of nonhomology. Protein sequence comparisons of the traM proteins from the F and R1-19 plasmids revealed several nonhomologous areas (Fig. V.3) which may be responsible for determining the plasmid specificity, although the majority of the differences are located within the first 40 residues.

Mullineaux and Willetts (21) have suggested that the product of the *finP* gene is a short RNA molecule that is capable of preventing transcription or translation or both of the *traJ* gene by interacting with its complementary RNA or DNA molecule, similar to the control of plasmid replication (27). The presence of a nonhomologous seven nucleotide loop flanked by homologous inverted repeats suggests that this loop could provide the *finP* plasmid specificity, possibly by allowing complementary base pairing at the loop when the *finP* gene is present, much like the RNA I interactions with the RNA primer in ColE1 plasmid replication (27).

The site of *finOP* action (*fisO*) has been reported to include both inverted repeats and the sequence immediately downstream (21). Because this area (except the loop) was conserved between the F and R1-19 plasmids, the *finO* product, which is relatively non-specific (34), probably recognizes this homologous area.

Although the -10 (nucleotide 1062) and -35 (nucleotide 1086) promoter regions of the R1-19 *finP* gene were highly homologous to those of the F plasmid (11 of 12 nucleotides) the intervening and surrounding sequences (in the amino terminal area of the *traJ* gene) were completely different. This is also true for the R100-1 plasmid (unpublished results). The significance of this promoter conservation is unclear at present.

The plasmid specificity of the *traJ*/P_{YZ} allele is closely linked to that of the *oriT*, *traY*, *traM* and *finP* alleles (34). When *traJ*/P_{YZ} sequences from the F (type I) and R1-19 (type III) plasmids were compared, no homology was found, although they are analogous systems. Both *traJ* products, when supplied *in trans*, increased transcription from their respective P_{YZ} promoters by similar levels, yielding similar amounts of highly homologous *traYZ* transcript. Because the F and R1-19 *traJ* products interacted with their P_{YZ} DNA, we examined both proteins for a possible DNA binding site by using the consensus sequence compiled by Ohlendorf *et al*. (22). However, we were unable to identify any predicted DNA binding domains in either protein sequence.

TABLE V.1

Galactokinase assay of F and R1-19 _traYZ_ promoter activity

Plasmid(s)	Promoter	_traJ_ product	GalK Units[a]
pKO4	None		63
pBF403	R1-19 P_{YZ}		175
pBF403/R1-19	R1-19 P_{YZ}	R1-19	547
pBF403/pBF405	R1-19 P_{YZ}	R1-19	658
pBF403/pRS27	R1-19 P_{YZ}	F	246
pBF302	F P_{YZ}		433
pBF302/pRS27	F P_{YZ}	F	807
pBF302/pBF405	F P_{YZ}	R1-19	373
pKL200	_lac_		646

[a] Nanomoles galactose phosphorylated per minute per unit of optical density at 650 nm.

FIGURE V.1.

Strategy for DNA sequencing of *traM*, *finP*, *traJ* and *traY* genes and the *traYZ* promoter from plasmid R1-19. Restriction sites for *Aha*III, *Bgl*II, *Pst*I, *Sau*3A, and *Taq*I are marked and the direction and length of sequenced DNA fragments are shown by arrows beneath the restriction map. The relative locations of the genes sequenced are also shown. [Sequence information for residues 25 to 474 was obtained from Ostermann *et al*. (23).]

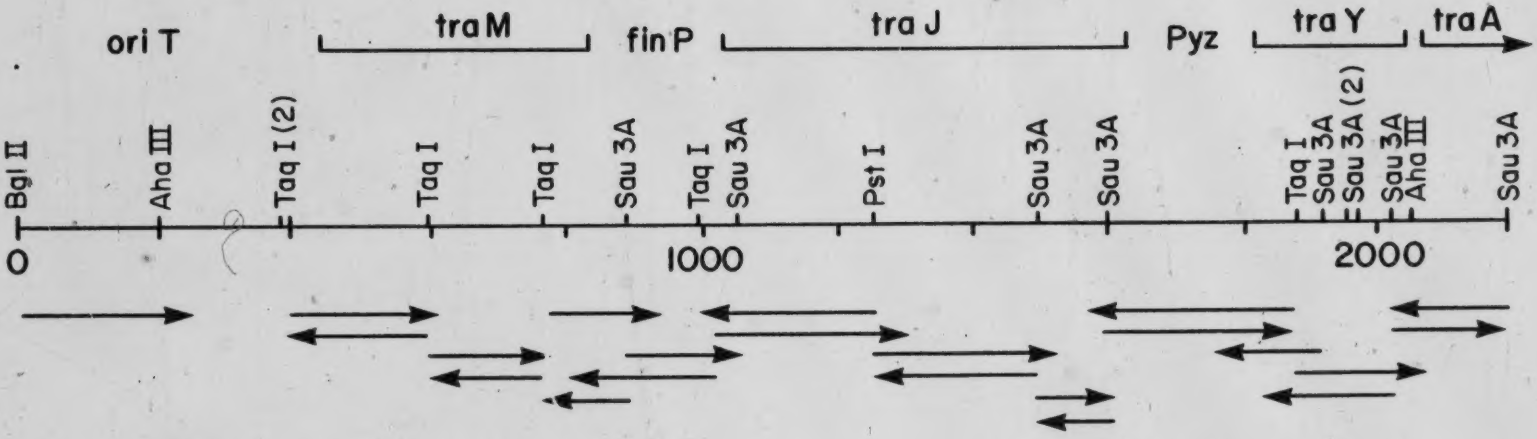

ori T
traM
fin P
tra J
Pyz
tra Y
tra A
Bgl II
Aha III
Taq I (2)
Taq I
Taq I
Sau 3A
Taq I
Sau 3A
Pst I
Sau 3A
Sau 3A
Taq I
Sau 3A
Sau 3A (2)
Sau 3A
Aha III
Sau 3A
0
1000
2000

FIGURE V.2

Nucleotide sequences of _oriT_, _traM_, _finP_, _traJ_ and _traY_ genes and the _traYZ_ promoter from plasmid R1-19. The sequence of nucleotides 25 to 474 (_oriT_) is from Ostermann _et_ _al_. (23), and the entire _traA_ sequence has been published previously (10). Potential -35 and -10 promoter regions for the _traM_, _traJ_, and _finP_ genes are boxed and inverted repeats are represented by arrows. Probable ribosome binding sites (29) are underlined. The _traYZ_ promoter was localized between the _traJ_ and _traY_ genes, but its exact location was not determined.

```
BglII                                                     60
AGATCTCATTTATAAACATCAGGCAGATGGCTAACATCCATTTTTTCATTTTTCCACCTC

                                                         120
TGGTGACTTTATCCGTAAATAATTTAACCCACTCCACAAAAAGGCTCAACAGGTTGGTGG

                                                         180
TTCTCACCACCAAAAGCACCACACCCCACGCAAAAACAAGTTTTTGCTGATTTGCTATTT

                                                         240
GAATCATTAACTTATGTTTTAAATAATGTATTTTAATTTATTTACATTATAAAAAGGAT

                                                         300
TCATTGGTGAATCGCATATGATTCACCAATGAATTTAACGCAACTCATAGATGAATCAGT

                                                         360
AATAATAAAACAACGAGTTAAAGCGAATTGACTCTAGATTCAATTCGTGATGTGATTTGC

                                                     -35 420
TGCATGAATCGTAATTTTCGTAACTTGCGTCACGATTCGATTCATCGATGAATTGTCATT

                -10                                 TraM 480
TTTTTATATGTGAAATAAATTTGATGCGATAAGAAAATAAAGAGGTATACATATGGCGAA
                                                    MetAlaLy

                                                         540
AGTGCAGGCTTATGTCAGTGACGAAATTGTCTATAAAATAAATAAAATTGTGGAAAGAAG
sValGlnAlaTyrValSerAspGluIleValTyrLysIleAsnLysIleValGluArgAr

                                                         600
ACGTGCAGAAGGAGCTAAAAGTACAGATGTAAGTTTCTCAAGCATATCCACTATGCTTCT
gArgAlaGluGlyAlaLysSerThrAspValSerPheSerSerIleSerThrMetLeuLe

                                                         660
CGAATTAGGGCTTCGTGTATATGAGGCTCAGATGGAGCGTAAAGAGTCAGCGTTCAATCA
uGluLeuGlyLeuArgValTyrGluAlaGlnMetGluArgLysGluSerAlaPheAsnGl

                                                         720
GGCTGAGTTTAATAAGGTACTTCTTGAATGTGCGGTAAAAACACAGTCAACAGTTGCTAA
nAlaGluPheAsnLysValLeuLeuGluCysAlaValLysThrGlnSerThrValAlaLy

                                                         780
GATTTTAGGTATTGAGTCACTAAGCCCTCATGTTTCCGGGAATCCAAAGTTTGAATATGC
sIleLeuGlyIleGluSerLeuSerProHisValSerGlyAsnProLysPheGluTyrAl

                                                         840
GAATATGGTCGAAGATATCAGGGATAAGGTGTCATCTGAGATGGAGAGGTTCTTTCCAGA
aAsnMetValGluAspIleArgAspLysValSerSerGluMetGluArgPhePheProGl

                                                         900
AAATGATGAGGAATAAAAAAGAATTAAAAGTTGATTTGCTTCAAATAACAGGATTTTGAT
uAsnAspGluGlu***

             -35                     -10                 960
CTGGTTCAATTTGATGACAATTAACCGCATAAAGGTTATATTAATTACGTGGTTAATGCC

                                       finP             1020
ACGTTAAAAATTGAAACTGAAAATCGCCGATGCAGGGAGGTCTGAACTCCCTGCATCGAC

                             TraJ           -10         1080
TGTCCATAGAATCCTTAACGGAGGTTCCTATGTGTGCGCTGGATCGTAGAGAAAGGCCAC
                             MetCysAlaLeuAspArgArgGluArgProL

     -35                                                1140
TTAACAGTCAATCTGTAAATAAATACATCCTTAACGTTCAGAATATCTACAGAAATTCTC
euAsnSerGlnSerValAsnLysTyrIleLeuAsnValGlnAsnIleTyrArgAsnSerP

                                                        1200
CCGTTCCGGTTTGTGTCCGTAACAAAAACCGGAAAATCCTTTATGCCAATGGGGCTTTTA
roValProValCysValArgAsnLysAsnArgLysIleLeuTyrAlaAsnGlyAlaPheI

                                                   PstI
TTGAACTCTTTTCCAGAGAAGATAAACCCTTATCCGGAGAGAGTTATATACGTCTGCAGG
leGluLeuPheSerArgGluAspLysProLeuSerGlyGluSerTyrIleArgLeuGlnV

                                                        1320
TTGAAATTTTTCTTTCATCACTTGAACTGGAATGCCAGGCTCTTGGACATGGCTCTGCAT
alGluIlePheLeuSerSerLeuGluLeuGluCysGlnAlaLeuGlyHisGlySerAlaP

                                                        1380
TTTGTCGTCGTTTTAATTTTCATGGCGAAATCTATCAGATAAGGATGGAGAATGTTTCTT
heCysArgArgPheAsnPheHisGlyGluIleTyrGlnIleArgMetGluAsnValSerP

                                                        1440
TTTATAATGACGAATCTGTTGTTTTATGGCAAATTAATCCGTTTCCTGATTATCCATTTT
heTyrAsnAspGluSerValValLeuTrpGlnIleAsnProPheProAspTyrProPheP

                                                        1500
TTGCGTTAAATCAGAGTGGAAGTAATACAAATACTTCTGATAAATTAACGATATGGAATG
heAlaLeuAsnGlnSerGlySerAsnThrAsnThrSerAspLysLeuThrIleTrpAsnA

                                                        1560
ATCTTTCTCCAGGGACATTGGTTGTTTTCTCTTTTTATATGCTGGGTGTTGGTCACGCAA
spLeuSerProGlyThrLeuValValPheSerPheTyrMetLeuGlyValGlyHisAlaT

                                                        1620
CAATTGCCAGAGAGTTGGGTATTACAGACAGAGCATCTGAGGATCGAATTAAACCAGTTA
hrIleAlaArgGluLeuGlyIleThrAspArgAlaSerGluAspArgIleLysProValL

                                                        1680
AACGGAAAATAAAAGAATTTTTTGAACACGTTTGATTTATTCAGAGTGTCATGTATCTAT
ysArgLysIleLysGluPhePheGluHisVal***

                                                        1740
AAAGGAGAAATAGATTCGCTATTAAGTATAATTCGTGAATTTTATGGTGTTAAGTAAATG

                                                        1800
TTAAATAAGGCGCGGCCAATTTTGGCGTGTTTTTTTTCGTCGGTTTCTGTGCAATCATGA

                               TraY                     1860
TTTTATTAGTCGGGAAACAGAGGTGAATTGAGTGAGGAGGCGTAACGCGAGAGGCGGGAT
                               MetArgArgArgAsnAlaArgGlyGlyIl

                                                        1920
AAGCAGAACAGTCTCAGTGTATCTCGATGAAGATACAAATAACCGACTTATTCGGGCCAA
eSerArgThrValSerValTyrLeuAspGluAspThrAsnAsnArgLeuIleArgAlaLy

                                                        1980
AGATCGTAGCGGTAGAAGTAAAACTATAGAGGTTCAGATCCGTCTCAGAGATCATCTGAA
sAspArgSerGlyArgSerLysThrIleGluValGlnIleArgLeuArgAspHisLeuLy

                                                        2040
ACGGTTTCCTGATTTCTATAATGAAGAAATATTCAGGGAGGTGATCGAAGAGAACGAGTC
sArgPheProAspPheTyrAsnGluGluIlePheArgGluValIleGluGluAsnGluSe

                                                    TraA2100
AACATTTAAAGAACTTTAGTTTTTATCTATCAATATTAAACAGAGGAATCATATGAATAC
rThrPheLysGluLeu***                                 MetAsnTh

                                                        2160
TGTTTTAAGTGTTCAGGGTGCTTCTGCGCCCGTCAAAAAGAAATCGTTTTTTTCTAAATT
rValLeuSerValGlnGlyAlaSerAlaProValLysLysLysSerPhePheSerLysPh

                                    2200
CACTCGTCTGAATATGCTTCGCCTGGCTCGCGCTGTGATC
eThrArgLeuAsnMetLeuArgLeuAlaArgAlaValIle
```

FIGURE V.3.

Comparison of plasmid F (first line) and plasmid R1-19 (second line) traM open reading frames. Nonhomologous residues are boxed. The traM sequence for plasmid F is from Thompson and Taylor (31).

```
                                        10                                       20
Met Ala Lys Val Asn Leu Tyr Ile Ser Asn Asp Ala Tyr Glu Lys Ile Asn Ala Ile Ile
Met Ala Lys Val Gln Ala Tyr Val Ser Asp Glu Ile Val Tyr Lys Ile Asn Lys Ile Val

                                        30                                       40
Glu Lys Arg Arg Gln Glu Gly Ala Arg Glu Lys Asp Val Ser Phe Ser Ala Thr Ala Ser
Glu Arg Arg Arg Ala Glu Gly Ala Lys Ser Thr Asp Val Ser Phe Ser Ser Ile Ser Thr

                                        50                                       60
Met Leu Leu Glu Leu Gly Leu Arg Val His Glu Ala Gln Met Glu Arg Lys Glu Ser Ala
Met Leu Leu Glu Leu Gly Leu Arg Val Tyr Glu Ala Gln Met Glu Arg Lys Glu Ser Ala

                                        70                                       80
Phe Asn Gln Thr Glu Phe Asn Lys Leu Leu Leu Glu Cys Val Val Lys Thr Gln Ser Ser
Phe Asn Gln Ala Glu Phe Asn Lys Val Leu Leu Glu Cys Ala Val Lys Thr Gln Ser Thr

                                        90                                      100
Val Ala Lys Ile Leu Gly Ile Glu Ser Leu Ser Pro His Val Ser Gly Asn Ser Lys Phe
Val Ala Lys Ile Leu Gly Ile Glu Ser Leu Ser Pro His Val Ser Gly Asn Pro Lys Phe

                                       110                                      120
Glu Tyr Ala Asn Met Val Glu Asp Ile Arg Glu Lys Val Ser Ser Glu Met Glu Arg Phe
Glu Tyr Ala Asn Met Val Glu Asp Ile Arg Asp Lys Val Ser Ser Glu Met Glu Arg Phe

                        127
Phe Pro Lys Asn Asp Asp Glu
Phe Pro Glu Asn Asp Glu Glu
```

FIGURE V.4.

Comparison of plasmid F (first line) and plasmid R1-19 (second line) _traY_ open reading frames. The _traY_ product from plasmid F is 44 amino acids larger than the equivalent product from plasmid R1-19. Nonhomologous residues are boxed, and spaces were inserted to allow maximum alignment. The _traY_ sequence for plasmid F is from Fowler _et al_. (9).

```
                                       10                                      20
Met Val Lys Leu Lys Leu Pro Val Asp Val Glu Ser Leu Leu Ile Glu Ala Ser Asn Arg

                                       30                                      40
Ser Gly Arg Ser Arg Ser Phe Glu Ala Val Ile Arg Leu Lys Asp His Leu His Arg Tyr

                                       50                                      60
Pro Lys Phe Asn Arg Ala Gly Asn Ile Tyr Gly Lys Ser Leu Val Lys Tyr Leu Thr Met
                Met Arg Arg Arg Asn Ala Arg Gly Gly Ile Ser Arg Thr Val Ser Val

                                       70                                      80
Arg Leu Asp Asp Glu Thr Asn Gln Leu Leu Ile Ala Ala Lys Asn Arg Ser Gly Trp Cys
Tyr Leu Asp Glu Asp Thr Asn Asn Arg Leu Ile Arg Ala Lys Asp Arg Ser Gly Arg Ser

                                   90                                         100
Lys Thr Asp Glu Ala Ala Asp Arg Val Ile Asp His Leu Ile Lys Phe Pro Asp Phe Tyr
Lys Thr Ile Glu Val.Gln Ile Arg Leu Arg Asp His Leu Lys Arg Phe Pro Asp Phe Tyr

                                      110                                     119
Asn Ser Glu Ile Phe Arg Glu Ala Asp Lys Glu Glu Asp Ile     Thr Phe Asn Thr Leu
Asn Glu Glu Ile Phe Arg Glu Val Ile     Glu Glu Asn Glu Ser Thr Phe Lys Glu Leu
```

E. BIBLIOGRAPHY

1. **Achtman, M., P. Manning, C. Edelbluth, and P. Herrlich.** 1979. Export without proteolytic processing of inner and outer membrane proteins encoded by F sex factor tra cistrons in Escherichia coli minicells. Proc. Natl. Acad. Sci. USA 76:4837-4841.

2. **Alfaro, G., and N. Willetts.** 1972. The relationship between the transfer systems of some bacterial plasmids. Genet. Res. 20:279-289.

3. **Birnboim, H.C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1523.

4. **Chang, A.C.Y., and S.N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. 134:1141-1156.

5. **Chou, P.Y., and G.D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. 47:45-148.

6. **Clark, A.J.** 1984. Conjugation and its aftereffects in E. coli. In Symposium on Origin and Evolution of Sex Marine. Biological Laboratory. Woods Hole, Mass. in press.

7. **Everett, R., and N. Willetts.** 1980. Characterization of an in vivo system for nicking at the origin of conjugal DNA transfer of the sex factor F. J. Mol. Biol. —

136:129-150.

8. **Finlay, B.B., L.S. Frost, and W. Paranchych**. 1984. Localization, cloning and sequence determination of the conjugative plasmid ColB2 pilin gene. J. Bacteriol. **160**:402-407.

9. **Fowler, T., L. Taylor, and R. Thompson**. 1983. The control region of the F plasmid transfer operon: DNA sequence of the traJ and traY genes and characterization of the traYZ promoter. Gene **26**:79-89.

10. **Frost, L.S., B.B. Finlay, A. Opgenorth, W. Paranchych and J. Lee**. 1985. Characterization and sequence analysis of pilin from F-like plasmids. J. Bacteriol. **164**:1238-1247.

11. **Frost, L.S., W. Paranchych, and N.S. Willetts**. 1984. DNA sequence of the F traALE region that incudes the gene for F pilin. J. Bacteriol. **160**:395-401.

12. **Gaffney, D., R. Skurray, and N. Willetts**. 1983. Regulation of the F conjugation genes studied by hybridization and tra-lacZ fusion. J. Mol. Biol. **168**:103-122.

13. **Hawley, D., and W. McClure**. 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res. 11:2237-2255.

14. **Hopp, T.P., and K.R. Woods**. 1981. Prediction of protein antigenic determinant from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**: 3824-3828.

15. **Humphreys, G., G.A. Willshaw, and E.S. Anderson.** 1975. A simple method for the preparation of large quantities of pure plasmid DNA. Biochim. Biophys. Acta **383**:457-463.

16. **Inouye, S., X. Soberon, T. Francheschini, K. Nakamura, K. Itakura, and M. Inouye.** 1982. Role of positive charge on the amino-terminal region of the signal peptide in protein secretion across the membrane. Proc. Natl. Acad. Sci. USA **79**:3438-3441.

17. **Koronakis, V. E., E. Bauer, and G. Hogenauer.** 1985. The traM gene of the resistance plasmid R1: comparison with the corresponding sequence of the Escherichia coli F factor. Gene **36**:79-86.

18. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1-45.

19. **McKenney, K, H. Shimatake, D. Court, U. Schmeissner, and M. Rosenberg.** 1981. A system to study promoter and terminator signals recognized by E. coli RNA polymerase. In J. Chirikjian and T. Papas, (eds.), Gene Amplification and Analysis, Vol.II, pp.383-415, Elsevier-North Holland, New York.

20. **Meynell, E., and N. Datta.** 1967. Mutant drug resistant factors of high transmissibility. Nature (London) **214**:885-887.

21. **Mullineaux, P., and N. Willetts.** 1984. Promoters in

the transfer region of plasmid F. In D. Helinski, S. Cohen, D. Clewell, and A. Hollaender (eds.), Plasmids in Bacteria, pp.605-614, Plenum Press, New York.

22. **Ohlendorf, D., W. Anderson, M. Lewis, C. Pabo and B. Matthews**. 1983. Comparison of the structures of cro and lamda repressor proteins from bacteriophage lambda. J. Mol. Biol. **169**:757-769.

23. **Ostermann, E., F. Kricek, and G. Hogenauer**. 1984. Cloning the origin of transfer region of the resistance plasmid R1. EMBO J. **3**:1731-1735.

24. **Raibaud, O. and M. Schwartz**. 1984. Positive control of transcription initiation in bacteria. Annu. Rev. Genet. **18**:173-206.

25. **Rosenberg, M., and D. Court**. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319-353.

26. **Sanger, F., S. Nicklen, and A.R. Coulson**. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463-5467.

27. **Scott, J.R.** 1984. Regulation of plasmid replication. Microbiol. Rev. **48**:1-23.

28. **Sharp, P.A., S.N. Cohen, and N. Davidson**. 1973. Electron microscope heteroduplex studies of sequence relations among plasmids of E. coli II. Studies of drug resistance (R) factors and F factors. J. Mol. Biol. **75**:235-255.

29. **Shine, J., and L. Dalgarno**. 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34-38.

30. **Skurray, R., H. Nagaishi, and A. Clark**. 1976. Molecular cloning of DNA from F sex factor of E. coli K-12. Proc. Natl. Acad. Sci. U.S.A. 73:64-68.

31. **Thompson, R., and L. Taylor**. 1982. Promoter mapping and DNA sequencing of the F plasmid transfer genes traM and traJ. Mol. Gen. Genet. **188**:513-518.

32. **Thompson, R., L. Taylor, K. Kelly, R. Everett, and N. Willetts**. 1984. The F plasmid origin of transfer: DNA sequence of wild-type and mutant origins and location of origin-specific nicks. EMBO J. **3**:1175-1180.

33. **Vieira, J., and J. Messing**. 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259-268.

34. **Willetts, N.S., and J. Maule**. 1985. Specificities of incF plasmid conjugation genes. Genet. Res. **47**:1-11.

35. **Willetts, N.S., and R. Skurray**. 1980. The conjugation system of F-like plasmids. Annu. Rev. Genet. **14**:41-76.

36. **Willetts, N., and B. Wilkins**. 1984. Processing of plasmid DNA during bacterial conjugation. Microbiol. Rev. **48**:24-41.

37. **Yanisch-Perron, C., J. Vieira, and J. Messing**. 1985. Improved M13 phage cloning vectors and host strains:

nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103-119.

Chapter VI

NUCLEOTIDE SEQUENCES OF FIVE IncF PLASMID *finP* ALLELES[1]

A. INTRODUCTION.

Conjugal DNA transfer requires expression of the transfer (*tra*) operon which is regulated in a two-stage process involving the *f*ertility *in*hibition system (*finO*, *finP*) and *traJ* (4,5). The products of *finO* and *finP* (FinO, FinP) are regulatory elements that act together at the *fisO* site (otherwise designated O_J or *traO*) to prevent *traJ* transcription. The *traJ* gene product, a positive control element, interacts with the P_{YZ} promoter, increasing transcription of the *traY*->*Z* operon (6,11,21).

All *finO* products from IncF plasmids studied to date are capable of complementing the natural *finO* mutation of the F plasmid (22) and inhibiting its transfer. There may be two types of *finO* products, differentiated by their varying abilities to inhibit this transfer (22). The *finO* genes from R100 and R6-5 have been cloned, but the *finO* product remains undefined (2,3,18).

[1] A version of this chapter has been accepted for publication. Finlay, B.B., L.S. Frost, W. Paranchych, and N.S. Willetts. J. Bacteriol.

finP is located between traM and traJ (8) and the nucleotide sequence of this region from F has been determined (17). Mullineaux and Willetts (11) suggested that finP is transcribed from a potential promoter within the translated region of the traJ transcript and that it is transcribed in the opposite direction to traJ transcription.

B. RESULTS AND DISCUSSION

FinP is plasmid-specific, since six finP alleles were found among twelve IncF plasmids studied (22). This genetic diversity as well as the lack of a suitable open reading frame in the finP transcript suggests that FinP may be an RNA molecule capable of interacting with its complementary DNA or RNA sequence as well as with FinO, preventing transcription and/or translation of traJ (11,22). Wild type finP regions of types II-V (22) from the IncF plasmids R124, ColVBtrp, R1-19, R100-1 and ColB4 were cloned and sequenced and compared to the sequence of the finP region of F (type I) (Figure VI.1). In addition, the sequences of three finP mutants [pED236 (ColB2Fdr, type II), pED200 (R124 finP, type II) and pED202 (R386 finP and/or fisO, type VI)] which lead to derepression of the transfer operon in these plasmids, were determined (Figure VI.1).

The finP genes of type I (F) and type II (ColVBtrp and R124) are highly homologous except for the striking difference in the region between the large inverted repeats (positions 97-103). The homology extends into the traM gene, with other single nucleotide differences occuring at positions 127, 211 and 243. The finP sequences of type III (R1-19), type IV (R100-1) and type V (ColB4) contain regions of conserved and altered sequence, compared to the type I and II sequences, suggesting that there is an evolutionary relationship between them. From positions 1 to 96 and 104 to 167, the type III and type V finP regions are identical, while the type IV sequence differs in several places. From position 167 to 258 type III and type IV sequences are highly homologous, whereas the type V sequence demonstrates more homology with the type I and type II sequences.

The finP promoter was localized within a BglII/TaqI fragment in F corresponding to positions 1 to 84 in Fig. VI.1 (11) and may overlap the BamHI site at position 35 in R100-1 (type IV) (Fee and Dempsey, manuscript submitted). It is evident that the -35 and -10 regions of a putative promoter are highly conserved in all the finP alleles examined (positions 10-38). Except for position 15, the -35 region of this finP promoter matches the prokaryotic consensus sequence TTGACa (7). The sequence at the proposed -10 region is conserved in four of the five alleles, with R100-1 differing at position 35. This sequence also strongly resembles the proposed consensus sequence for the -10 region of a prokaryotic promoter [TAtaAT; (7)]. The sequence of the finP mutation in pED236 supports the concept that this is the finP promoter (Fig. VI.1) since a single nucleotide substitution at position 38 (T->C) could be interpreted as a deleterious mutation in the -10 region of the finP promoter.

Previously the finP type of R386 was assigned to a separate group (type VI), but it should be noted that genetic determination of the relationship between F and R386 finP specificities is rendered difficult by their incompatibility, and by the high background level of R386 transfer (22). However, the sequence of pED202 (R386 finP and/or fisO) suggests that it has a type I finP allele since it differs from F only at position 34; the observed nucleotide difference could be a mutation in the finP promoter resulting in derepression of the transfer operon. Confirmation of this requires sequencing of the finP region of wild type R386.

Comparison of the five finP sequences located between an inverted repeat (positions 61 to 66, Fig. VI.1) revealed several differences. The sequences of types I and II finP genes are identical in this region, as are types III and V. Type IV encodes a sequence similar to types I and II, differing only at position 64. These changes do not match the corresponding finP alleles reported by Willetts and

Maule (22), indicating that this region probably does not determine _finP_ specificity.

The most striking differences between the five _finP_ genes are located in the loop between the 12 bp inverted repeats (positions 97-103, Fig. VI.1 and VI.2), where all five _finP_ genes have unique sequences. The nucleotide sequence of type III is related to type V by simply inverting the seven base pair loop. Four of seven base pairs in the loops of type I and II are also related in this way. The original suggestion of Johnson _et_ _al_. (8) that FinP is a peptide could not be entirely excluded by Mullineaux and Willetts (11) or Fee and Dempsey (manuscript submitted). However, the sequence of these _finP_ alleles argues strongly against this since R100-1 (type IV) has a six base pair loop resulting in a frame shift mutation in the putative _finP_ peptide and ColB4 (type V) contains a stop codon within its loop which would give a truncated peptide if translation begins at position 76. We conclude that the product of _finP_ is indeed an RNA molecule.

The 12 bp inverted repeat (positions 85-115, Fig. VI.1) is followed by a sequence of four T's (positions 120-123). It is possible that this is a transcriptional terminator sequence (13) and signifies the end of the _finP_ transcriptional unit, as proposed by Mullineaux and Willetts (11). However, Fee and Dempsey have recently reported the presence of two _finP_ RNA molecules of 100 and

150 bp, using the R100 finP system (manuscript submitted), suggesting that transcriptional termination may occur near positions 146 and 196 in Fig. VI.1.

The concept of FinP being an RNA molecule is attractive, since anti-sense RNA has been shown to be a control element in a number of prokaryotic systems. Examples are found within the ColE1 plasmid replication system (14), R1 plasmid replication (15) and the ompF gene (10). For the control of plasmid replication, antisense RNA forms a stem-and-loop structure resembling that of tRNA where one of the loops (analogous to the anti-codon loop) is plasmid-specific, providing the basis for incompatibility.

When the finP sequences were examined for regions capable of forming secondary structure, several were found. The sequence that showed the most potential for stem-and-loop formation was the large inverted repeat surrounding the finP-specific region (positions 97-103). A second possible stem-and-loop structure could encompass positions 49-78. A model of the finP RNA secondary structure is shown in Figure VI.2. The stems of the putative stem-and-loop structures 1 and 2 are highly conserved, suggesting that the secondary structure of the finP RNA molecule may be important. The G to A transition at position 70 in pED200 (R124 finP) decreases the predicted free energy of stem-and-loop structure 1 from

-6.2 Kcal/mol to -0.4 Kcal/mol destroying its stability and lending further support to the idea that FinP is an RNA molecule. The two F finP mutations reported by Mullineaux and Willetts (11) are not as easily explained by this RNA model. The mutation at position 78 within stem 1 (pED234) decreases the predicted free energy from -6.2 to -4.0 Kcal/mol while the second mutant (pED235) does not affect this structure. However, these mutations occur within a completely conserved region and may be involved in interactions which are not yet understood.

If the finO product is involved in an interaction with FinP, this interaction would presumably occur in a highly conserved region of finP. Examination of Figures VI.1 and VI.2 reveals several such regions. These include the stems of stem-and-loop structures 1, 2 and the region between them (Fig. 2) with the loop of stem-and-loop structure 1 possibly determining finO specificity as discerned by complementation of the finO mutation of F (22). It is possible that FinO may interact with one or more of these sequences to form the active FinOP repressor. This may alter the finP RNA molecule by affecting its processing, or allow the finP RNA molecule to adopt a conformation necessary for interaction with traJ or its transcript. The two F finP mutations (pED234, pED235) may be within such a recognition sequence, producing the FinP$^-$ phenotype by preventing the FinO/FinP interaction.

FIGURE VI.1

Nucleotide sequence of the finP regions of plasmids with type I (F), type II (R124 and ColVBtrp, which were identical), type III (R1-19), type IV (R100-1) and type V (ColB4) alleles, of two finP mutations [pED236 (ColB2Fdr, type II) and pED200 (R124 finP, type II)] and of pED202 (R386 finP and/or fisO, type VI). The sequence of the two F finP mutations in pED234 and pED235 [derived by hydroxylamine mutagenesis of pED812; (8)] at positions 78 and 79 respectively are from Mullineaux and Willetts (11). All IncF plasmids used in this study are described by Willetts and Maule (22). Nucleotides that are not homologous with the type I finP sequence are boxed. The location of the putative finP promoter and transcriptional start site are taken from Mullineaux and Willetts (11) while the proposed traJ promoter, the traJ transcriptional start site, the traJ translational start site (TraJp) and the traM translational stop site (traMp) which are encoded by the complementary strand are taken from Thompson and Taylor (17). Also shown are the three inverted repeats in this region, the third set being the proposed terminator for traM transcription (17). BglII digests of plasmids encoding type I or type II finP genes were ligated into pUC8 (19) and chimera-containing colonies were immobilized on nitrocellulose and probed with a ^{32}P-labelled DNA probe (9) made from the 1.1 kb BglII fragment of F that contains oriT, traM, and finP (16,17). Plasmid DNA was prepared (1) from colonies containing the 1.1 kb BglII fragment and sequenced by the method of Wallace et al. (20) using reverse or 17-bp primers (New England Biolabs). The finP region of R100-1 was obtained by subcloning the 1.0 kb BamHI fragment and the 2.4 kb BamHI/HindIII fragment that contain finP and sequenced in a similar manner. The nucleotide sequences of types III and V finP genes were determined by the strategy described previously (Chapter V).

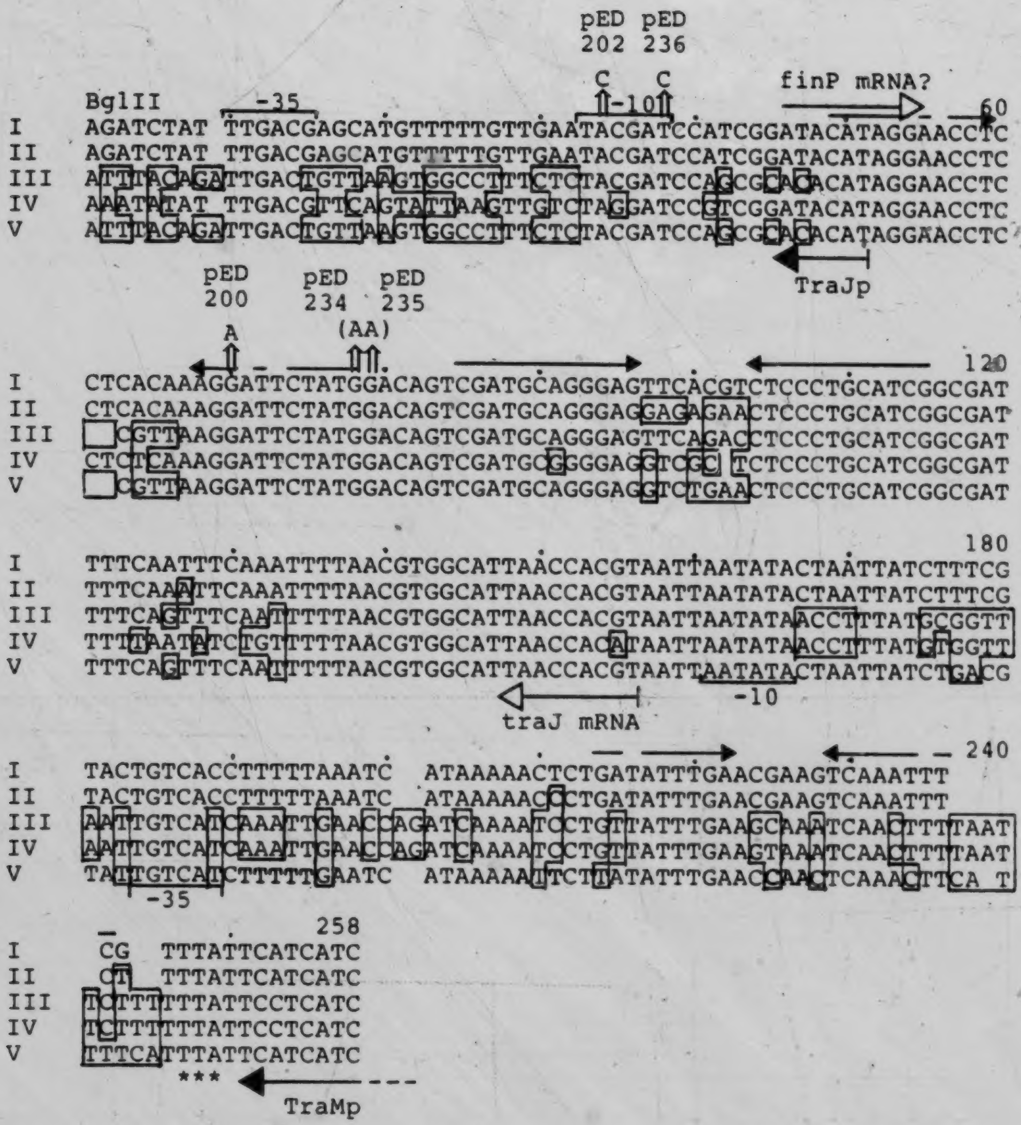
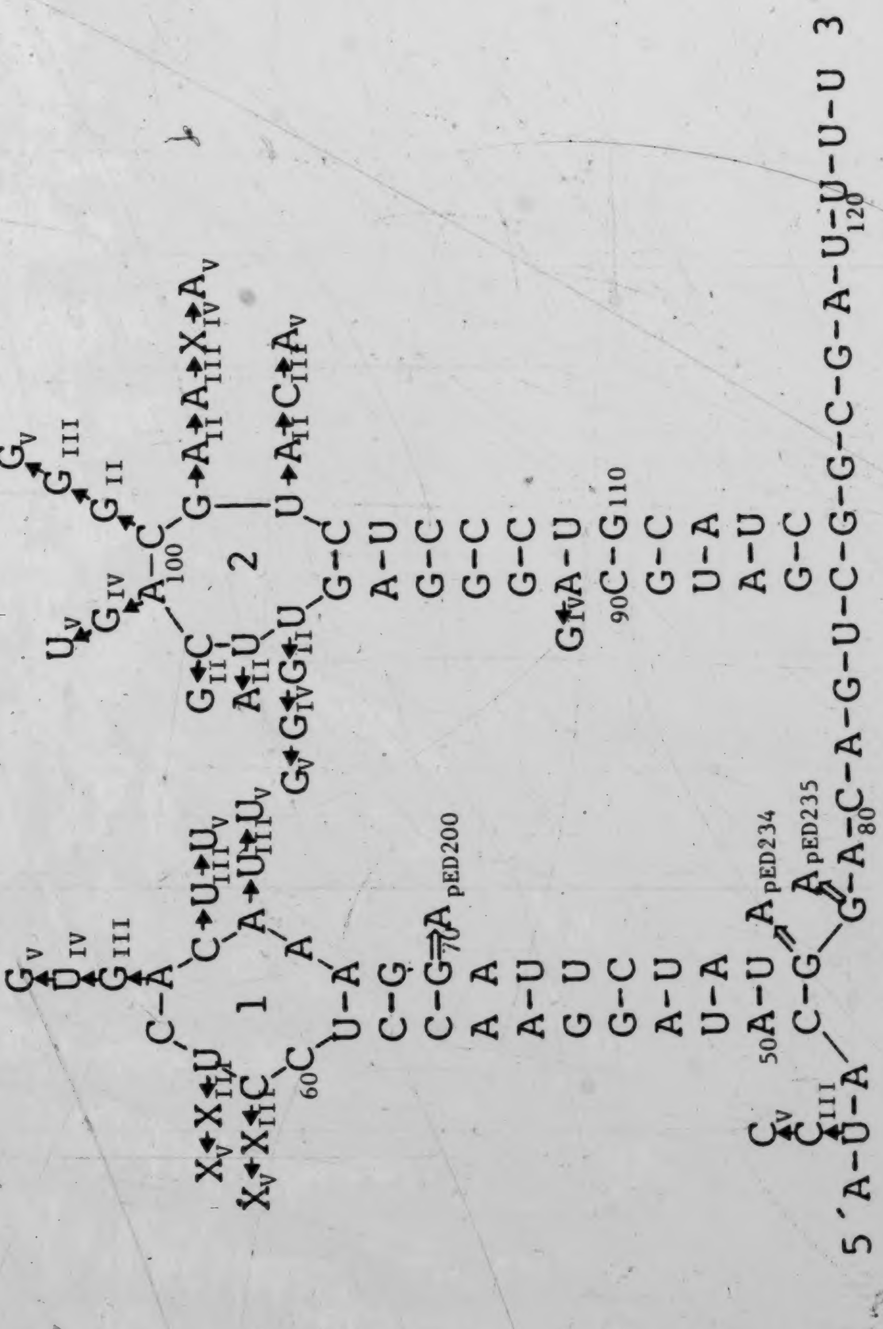

pED 202
pED 236
C
C
finP mRNA?
BglII
-35
-10
60
I AGATCTAT TTGACGAGCATGTTTTTGTTGAATACGATCCATCGGATACATAGGAACCTC
II AGATCTAT TTGACGAGCATGTTTTTGTTGAATACGATCCATCGGATACATAGGAACCTC
III ATTTACAGATTGACTGTTAAGTGGCCTTTCTCTACGATCCAGCGCACACATAGGAACCTC
IV AAATATAT TTGACGTTCAGTATTAAGTTGTCTAGGATCCGTCGGATACATAGGAACCTC
V ATTTACAGATTGACTGTTAAGTGGCCTTTCTCTACGATCCAGCGCACACATAGGAACCTC
TraJp
pED 200
pED 234
pED 235
A
(AA)
120
I CTCACAAAGGATTCTATGGACAGTCGATGCAGGGAGTTCACGTCTCCCTGCATCGGCGAT
II CTCACAAAGGATTCTATGGACAGTCGATGCAGGGAGGAGAGAACTCCCTGCATCGGCGAT
III CGTTAAGGATTCTATGGACAGTCGATGCAGGGAGTTCAGACCTCCCTGCATCGGCGAT
IV CTCTCAAAGGATTCTATGGACAGTCGATGCGGGGAGGTCGC TCTCCCTGCATCGGCGAT
V CGTTAAGGATTCTATGGACAGTCGATGCAGGGAGGTCTGAACTCCCTGCATCGGCGAT
180
I TTTCAATTTCAAATTTTAACGTGGCATTAACCACGTAATTAATATACTAATTATCTTTCG
II TTTCAAATTCAAATTTTAACGTGGCATTAACCACGTAATTAATATACTAATTATCTTTCG
III TTTCAGTTTCAATTTTTAACGTGGCATTAACCACGTAATTAATATAACCTTTATGCGGTT
IV TTTTAATATCTGTTTTTAACGTGGCATTAACCACATAATTAATATAACCTTTATGTGGTT
V TTTCAGTTTCAATTTTTAACGTGGCATTAACCACGTAATTAATATACTAATTATCTGACG
-10
traJ mRNA
240
I TACTGTCACCTTTTTAAATC ATAAAAACTCTGATATTTGAACGAAGTCAAATTT
II TACTGTCACCTTTTTAAATC ATAAAAACCCTGATATTTGAACGAAGTCAAATTT
III AATTGTCATCAAATTGAACCAGATCAAAATCCTGTTATTTGAAGCAAATCAACTTTTAAT
IV AATTGTCATCAAATTGAACCAGATCAAAATCCTGTTATTTGAAGTAAATCAACTTTTAAT
V TATTGTCATCTTTTTGAATC ATAAAAATTCTTATATTTGAACCAACTCAAACTTCA T
-35
258
I CG TTTATTCATCATC
II CT TTTATTCATCATC
III TCTTTTTTATTCCTCATC
IV TCTTTTTTATTCCTCATC
V TTTCATTTATTCATCATC
***
TraMp

FIGURE VI.2

Possible secondary structure of a type I finP RNA molecule corresponding to positions 46-123, Fig. VI.1. The two stem-loop structures have predicted free energies of -6.2 Kcal (structure 1) and -24.2 Kcal (structure 2) (12). Loop 2 determines finP specificity (see text). Nucleotides are numbered as in Fig. VI.1. Roman numeral subscripts represent nucleotides from types II-V finP genes that differ when compared to type I and an X represents a space inserted to allow maximum alignment.


pED200
pED234
pED235

C. BIBLIOGRAPHY

1. **Birnboim, H.C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1523.

2. **Cheah, K., A. Ray and R. Skurray.** 1984. Cloning and molecular analysis of the finO region from the antibiotic-resistance plasmid R6-5. Plasmid 12: 222-226.

3. **Dempsey, W.B. and S.A. McIntire.** 1983. The finO gene of antibiotic resistance plasmid R100. Mol. Gen. Genet. 190:444-451.

4. **Finnegan, D.J., and N.S. Willetts.** 1971. Two classes of Flac mutants insensitive to transfer inhibition by an F-like factor. Mol. Gen. Genet. 111:256-264.

5. **Finnegan, D.J., and N.S. Willetts.** 1973. The site of action of the F transfer inhibitor. Mol. Gen. Genet. 127:307-316.

6. **Gaffney, D., R. Skurray, and N. Willetts.** 1983. Regulation of the F conjugation genes studied by hybridization and tra-lacZ fusion. J. Mol. Biol. 168:103-122.

7. **Hawley, D., and W. McClure.** 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res. 11:2237-2255.

8. **Johnson, D., Everett, R. and N. Willetts.** 1981. Cloning of F DNA fragments carrying the origin of transfer

oriT and the fertility inhibition gene finP. J. Mol. Biol. 153:187-202.

9. **Maniatis, T., E.F. Fritsch, and J. Sambrook**. 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

10. **Mizuno, T., M. Chow, and M. Inouye**. 1984. A unique mechanism regulating gene expression: Translational inhibition by a complementary RNA transcript (mic RNA). Proc. Natl. Acad. Sci. USA 81:1966-1970.

11. **Mullineaux, P., and N. Willetts**. 1984. Promoters in the transfer region of plasmid F. In D. Helinski, S. Cohen, D. Clewell, and A. Hollaender (eds.), Plasmids in Bacteria, pp.605-614, Plenum Press, New York.

12. **Queen, C., and L.J. Korn**. 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids. Res. 12:581-599.

13. **Rosenberg, M. and D. Court**. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. 13:319-353.

14. **Scott, J.R.** 1984. Regulation of plasmid replication. Microbiol. Rev. 48:1-23.

15. **Stougaard, P., S. Molin, and K. Nordstrom**. 1981. RNAs involved in copy-number control and incompatibility of plasmid R1. Proc. Natl. Acad. Sci. USA 78:6008-6012.

16. **Thompson, R. and M. Achtman**. 1978. The control region of the F sex factor DNA transfer cistrons:

Restriciton mapping and DNA cloning. Mol. Gen. Genet. 165:295-304.

17. Thompson, R., and L. Taylor. 1982. Promoter mapping and DNA sequencing of the F plasmid transfer genes traM and traJ. Mol. Gen. Genet. 188:513-518.

18. Timmis, K., I. Andres, and M. Achtman. 1978. Fertility repression of F-like conjugative plasmids: Physical mapping of the R6-5 finO and finP cistrons and identification of the finO protein. Proc. Natl. Acad. Sci. USA 75:5836-5840.

19. Vieira, J., and J. Messing. 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259-268.

20. Wallace, R.B., M.J. Johnson, S.V. Suggs, K. Miyoshi, R. Bhatt, and K. Itakura. 1981. A set of synthetic oligodeoxyribonucleotide primers for DNA sequencing in the plasmids vector pBR322. Gene 16:21-26.

21. Willetts, N.S. 1977. The transcriptional control of fertility in F-like plasmids. J. Mol. Biol. 112:141-148.

22. Willetts, N.S., and J. Maule. 1985. Specificities of incF plasmid conjugation genes. Genet. Res. 47:1-11.

Chapter VII

STUDIES ON THE ORIGIN OF TRANSFER OF IncF PLASMIDS: NUCLEOTIDE SEQUENCES OF THE TYPE II oriT/traM/traY ALLELE FROM ColB4-K98 AND THE TYPE IV traY ALLELE FROM R100-1[1]

A. INTRODUCTION

Bacterial conjugation is the process of transferring plasmid DNA inter- or intragenerically and is encoded by the transfer (tra) region of the bacterial plasmid. The tra operons of plasmids belonging to the IncF incompatibility groups share a high degree of DNA homology within this region (20). The transfer genes of the F plasmid are the most extensively studied of the IncF plasmids (for a review see Willetts and Skurray [28]), and most of these tra genes are encoded within the 33 kilobase (kb) traYZ operon. Genes of the transfer region of F (and related IncF plasmids) can be divided into five main groups, based on their functions: pilus biosynthesis (traALEKBVWCUFQHG), stabilization of mating pairs (traNG), surface exclusion (traST), conjugal DNA metabolism (oriT, traMYDIZ) and regulation (finOP, traJ and P_{YZ} promoter)

---

[1] A version of this chapter has been submitted.
Finlay, B.B., L.S. Frost, and W. Paranchych.
J. Bacteriol.

(28). Although the *tra* regions of IncF plasmids demonstrate homology, they are not identical. Several alleles of plasmid-specific *tra* genes have been identified (27). One of these allele groups includes *oriT*, *traM* and *traY* (27, 28). Willetts and Maule identified four *oriT*/*traM*/*traY* alleles from twelve IncF plasmids tested (27): type I (F, ColV2, R386, ColB2, R124, and ColVB*trp*), type II (ColB4-K98 and R538-1), type III (R1) and type IV (R6-5, R100, R136).

During plasmid-mediated conjugation, a single strand of plasmid DNA is transferred from the donor cell to the recipient cell, initiating at the origin of transfer (*oriT*). *OriT* defines the region required for the binding of proteins involved in nicking at *oriT* and triggering conjugal DNA metabolism such that a single strand is transferred in the 5'->3' direction with the region upstream of *oriT* passing into the recipient cell first. Consequently the transfer region is the last portion of the plasmid to be transferred (for a review, see Willetts and Wilkins [29]). Thompson *et al*. have sequenced F *oriT* and identified a cluster of nick sites by examining the strand-specific nicks acquired by a *lambda*-*oriT* genome in the presence of the F plasmid (22). The *oriT* nucleotide sequence from the R1 plasmid has also been reported (15).

Everett and Willetts (4) demonstrated that the products of *traY* and *traZ* are required for nicking at F oriT in vivo. Several oriT⁻ mutations have been sequenced and were found to alter one of two positions near the cluster of nick sites within oriT (22,29). Willetts and Maule proposed that the plasmid-specific traY product (TraYp) recognizes a DNA sequence in the oriT region while the traZ protein interacts with TraYp to form the traYZ endonuclease which is responsible for the nicking action (27). The nucleotide sequence of traY from F (8) and R1 (7) has been published, and it was shown that these traY products share some homology (7).

In F, the traM gene product is thought to recognize a plasmid-specific oriT sequence (27,29) and cause triggering of conjugal DNA synthesis (12,29). The traM gene has been sequenced from F (21), R1 (7,13) and R100 (5) plasmids.

We report here the nucleotide sequence of the type II traM/traY/oriT allele from pED203, a finO⁻ derivative of ColB4-K98 (27) (hereafter referred to as ColB4), and traY of R100-1, a type IV allele, and compare these sequences with the related genes from other IncF plasmids.

B. MATERIALS AND METHODS

1.Bacterial strains and plasmids.

pED203 and R100-1, the finO⁻ derivatives of ColB4-K98 and R100 (27), are IncFII plasmids which were carried in the host strain Escherichia coli JC6256 (F⁻

lac trp). pBF701 is a recombinant containing the 1.2 kb *Bgl*II/*Pst*I fragment from pED203 ligated into pUC8. pAOB4 is the 2.1 kb *Pst*I fragment from pED203 (which contains *traYALE* [9]) inserted into pUC8. pLF100 is a 3.7 kb *Bam*HI-*Pst*I fragment containing a truncated *traJ* gene and *traYALE* (9). All pUC8 recombinants were transformed into JM83 (24).

2. **Plasmid purification.** All plasmids were isolated and purified by the method of Humphries *et al*. (11) while small amounts of chimeric plasmids to be used for sequencing were obtained by the method of Birnboim and Doly (2).

3. **Recombinant DNA techniques.** These were performed as described previously (6,7,9).

4. **DNA sequencing.** Plasmid DNA was sequenced by the method of Wallace *et al*. (25) using the dideoxy sequencing method (19). Reverse and 17-bp primers were obtained from New England Biolabs (MA).

C. RESULTS

1. Cloning and Sequencing Strategy

pED203 (ColB4-K98 finO$^-$) plasmid DNA was digested with *Bgl*II and *Pst*I and electrophoresed on a 0.7% agarose gel, transferred to nitrocellulose and probed with the 1.2 Kb BglII/PstI fragment from R1-19 (which contains oriT, traM, finP and part of traJ [7]). A 1.2 Kb BglII/PstI fragment from ColB4, showed homology (data not shown) and was subsequently cloned into pUC8 and the recombinant called pBF701 (Fig. VII.1). The cloning and partial sequence of the 2.1 Kb PstI fragment from pED203 which contains traYALE and part of traJ (pAOB4) has been described elsewhere (9).

The sequence of traY from R100-1 was determined from the region immediately upstream of traA using the chimera pLF100 which contains a 3.7 Kb BamHI/PstI fragment and pUC8 (9).

Subclones of pLF100, pBF701 and pAOB4 were constructed in pUC8 and sequenced by the dideoxy sequencing technique (19,25) using linearized DNA isolated by the rapid alkali method (2). Fig. VII.1 outlines the regions sequenced from ColB4, and the relative positions of the various tra genes encoded by pBF701 and part of pAOB4. The position of the traY gene in pLF100 is equivalent to that for ColB4 and the sequence of the R100 traJ gene and the promoter for the traYZ transcript (P_{YZ}) are the subject of another paper (Frost, Finlay, Paranchych, in preparation). The sequence of the traA pilin genes for ColB4 and R100-1 (9) has been presented previously as has the sequence encoding finP (Chapter VI). When the partial nucleotide sequence of ColB4 *traJ* was determined, it was found to be identical to the R1 *traJ* sequence (7, data not shown), while the partial sequence of the ColB4 P_{YZ} promoter region differed from the equivalent R1 sequence at three positions (Fig. VII.2B). This similarity between R1 and ColB4 *traJ* and P_{YZ} promoters was expected since they belong to the same *traJ*/P_{YZ} allele group (type III) and are thought to be interchangable (27). For these reasons the remainder of *traJ* and P_{YZ} were not sequenced. Figure VII.2 contains the entire nucleotide sequence of *oriT*, *traM* and *traY*, and the predicted translation products of *traM* and *traY* from pED203. It also shows the nucleotide sequence and predicted amino acid sequence for *traY* of R100-1.

2. *OriT*.

The sequence of the origins of transfer of the F plasmid (22), R100 plasmid (W. Dempsey, personal communication) and the R1 plasmid (15) have been established. When the ColB4 *oriT* sequence is aligned with these sequences, the first 173 nucleotides are identical except for position 22 in ColB4 and positions 6, 29, 48, 145, 147 and 171 in R100 (Fig. VII.3, Table VII.1). This region includes the *oriT* nick sites and the two inverted repeats found near these sites (Figs. VII.2A, VII.3) as well as an open reading frame extending to position 170.

The promoter and ribosome binding site for gene X (22), encoded by the opposite strand to that presented here, are conserved (positions 109 to 56, Fig. VII.3), and will be discussed below.

From position 174 to the beginning of traM translation, there is limited DNA homology (65-80%) when aligned by the program of Queen and Korn (17). However, three distinct regions can be discerned within this sequence (Fig. VII.3, Table VII.I). There is an adenine-thymine (AT) rich region (positions 152-237) where ColB4 and R1 are nearly identical (96% homology) and F and R100 are different. Within a second region (positions 232-391), ColB4 is more closely related to F and R100 than to R1 (Table VII.1). While there is homology between the various plasmids in this area, each plasmid has unique features in terms of possible secondary structure. The possible significance of these structures will be discussed below. The third region (positions 392-464) immediately precedes the traM gene and is thought to contain transcriptional promoter(s) and a ribosome binding site (RBS) required for traM expression.

Although the 85 nucleotide AT rich sequence of ColB4 (positions 152-237) is not highly homologous to R100 or F, it has maintained a high level of A-T content in this region (84% compared to 80% in F). This A-T rich sequence is highly homologous to the AT-rich region in R1, including a 10 bp inverted repeat (IVR) (postions 178-203) which closely resembles the 12 bp IVR in R1. F and R100 have unique IVRs in comparable locations.

In the F and R1 plasmids, a large inverted repeat immediately follows this AT-rich sequence (positions 233-280, Fig. VII.3). The predicted free energy of formation for this structure in single-stranded DNA is -31.6 Kcal for F and -20.4 Kcal for R1. When the equivalent region from ColB4 is examined, a 10 nucleotide inverted repeat is found (positions 235-256 in Fig. VII.2A, VII.3) which has a predicted free energy of -18.4 Kcal. It is interesting to note that the sequence 5'-TAGGGTCG-3' is present twice in the upper strand of ColB4 starting at positions 236 and 280, and is repeated three times in succession in the bottom strand at positions 277, 266 and 255 (with a total of 2 mismatches). R100 contains two inverted repeats in this region: one between positions 207-237 (Fig. VII.3); and one between positions 232-280.

The sequence from position 287 to 390 in ColB4 is similar to the equivalent region in F and R100-1 (Figure VII.3) in that it contains two large direct repeats (positions 324-344 and 370-390) which have high homology with the equivalent 22 bp repeats found in F. However, this region also contains a large inverted repeat, encompassing positions 332-345 and 373-387 (-39.8 Kcal) which is unique to ColB4. The sequence 5'-TAGGGTCG-3' is repeated five times in this region, with three repeats having one mismatch and one having two mismatches. These repeats begin at positions 324, 370, 382 in the upper strand and at 337 and 371 in the lower strand (Figure VII.3). In all, this sequence is repeated ten times within positions 236-389. This short sequence is not found in the other three plasmids nor is there a pattern of short direct repeats of differing sequence. While F, R100-1 and ColB4 resemble each other in this region in terms of homology (68-81%, Table VII.1) and direct repeats (in the 320-390 region), R1 is nonhomologous and contains one 14 bp inverted repeat (position 322-348).

The location of the promoters for traM transcription have been proposed for F (positions 332-364; [21]) and for R1 (positions 349-377 and 413-441;[13]) but neither of the two R1 traM promoters nor the putative F traM promoter (Figure VII.3) are conserved in ColB4. Examining the ColB4 sequence for regions that match the consensus sequence for procaryotic promoters (10) revealed two possible promoter sequences (positions 291-319 and 400-427; Figure VII.2A and VII.3). The putative traM promoters in R100 (Fig. VII.3) were chosen for their agreement with the promoter consensus sequence and for their high homology to the equivalent promoters in ColB4. Confirmation of these transcriptional initiation sites awaits sizing of RNA transcripts in this region. All of these traM genes contain a conserved ribosome binding site (AGGT). The ColB4, R100 and F plasmids contain eight non-conserved nucleotides between this sequence and the putative traM translational start codon while R1 contains six nucleotides between these two sequences.

3. The TraM protein

When the nucleotide sequence presented in Fig. VII.2A was examined for possible translational products, a 127 residue polypeptide (positions 464-844, Fig. VII.2A) was found. This polypeptide, which we consider to be the ColB4 traM gene product, was highly homologous to the predicted traM products from F, R1 and R100 (see Table VII.1, Fig. VII.4). The majority of differences between these proteins reside within the first 40 residues (Figure 4) and differences found after residue 40 are for the most part conservative amino acid substitutions.

Originating 13 nucleotides downstream from the traM translational stop site is a small (5 bp) region of dyad symmetry (positions 860-875, Fig. VII.2A). F, R100 and R1 contain an equivalent sequence downstream of traM (5,7,13,21). It has been suggested that these unconserved inverted repeats are the traM transcriptional terminators. This sequence from ColB4 may be the traM terminator, terminating transcription upstream of the traJ promoter. Since this putative transcriptional terminator bears little resemblance to a rho-independent terminator consensus sequence (18), it is probably dependent upon the rho termination factor.

4. The traY protein

The nucleotide sequences of the regions encoding traY for ColB4 and R100-1 were determined and are shown in Fig. VII.2. The DNA sequence for ColB4 traY is highly homologous to the equivalent sequence from R1 (7). There are 13 nucleotide differences found between these two traY sequences, which translate into three amino acid differences (Fig. VII.2B). The proposed translational start codon of the ColB4 traY product is GUG, as is found in R1 (Table VII.1;[7]), with the preceeding ribosome binding site completely conserved between these two plasmids. The intercistronic region between traY and traA is completely conserved but amino acid residues 2 and 3 of the traA product are different between R1 and ColB4. The remainder of the ColB4 and R1 pilin sequences have been published elsewhere (9). The DNA sequence for R100-1 traY shares homology with F and ColB4/R1 traY (57-60%, Table VII.1) while the translated protein demonstrates limited homology (39-49%) with the corresponding traY proteins (Figure VII.5). Disregarding conservative substitutions, the majority of the changes are localized in the amino terminal portion of the protein with other substitutions distributed randomly, especially near the carboxy terminus. Interestingly, the F traY protein has an additional 44 amino acids at its amino terminus. The protein encoded by R100-1 traY begins with an unusual initiator codon (14), UUG, which is reminiscent of ColB4 and R1 traY (GUG). These three traY genes all contain strong ribosome-binding sites which may help to offset the rare initiator codon usage in these genes.

D. DISCUSSION

The sequences presented here complete our knowledge at the DNA level of the four traM and oriT alleles and three traY alleles of IncF plasmids. Other transfer genes involved in DNA metabolism during conjugation which are thought to act in the oriT region include traIDZ which are encoded at the end of the transfer operon, distal to oriT. There are three traI groups represented by F, R100 and R1/ColB4 where the R1/ColB4 traI gene products are similar to the F traI product but act less efficiently at complementing a traI mutation (27). The traZ gene products from F and R100 were shown to be nonequivalent while the R1 and F traZ products are interchangeable (4). The specificity of the ColB4 traZ allele has not been determined.

Several interesting observations can be made by comparing the DNA sequence of the oriT regions from F, R100, R1 and ColB4. It has been suggested that the traZ component of the traYZ endonuclease may be responsible for the nicking activity at oriT (27). The sequence at the nicking sites presented by Thompson et al. (22) is completely conserved in all four plasmids, suggesting that the domain of the traZ product which is responsible for the endonuclease activity should be conserved. The specificity of traZ action noted above may be defined either by the few differences in the conserved region of oriT (positions 145, 147, 171) or, more likely, by the proposed interaction of traZ product with the traY protein, which may provide the DNA binding function of the traYZ endonuclease (27). This model also suggests that one domain of the traY product would bind a plasmid-specific DNA sequence (one of three alleles) in the oriT region while a second domain would interact with the traZ product (one of two alleles) and align it correctly with the DNA, allowing nicking at oriT.

The 85 nucleotide A-T rich sequence (positions 152-237) is nearly identical in ColB4 and R1, but is not conserved in F or R100, although the equivalent regions in F and R100 are also A-T rich. While this A-T rich region could aid in initiating strand separation (a requirement of traI activity) as suggested by Willetts and Wilkins (29), we propose that it acts as a recognition site for the traY component of the traYZ endonuclease. The amino terminal region of the various traY products (Fig. VII.5) show the greatest variability in protein sequence and can be divided into three groups; F, R1/ColB4 and R100. This domain of the protein could be responsible for recognition of the three unique DNA sequences (F, R1/ColB4, R100) in the AT-rich region of oriT. Other domains of the protein which could be involved in binding to the traZ product are not as readily discernible.

Two large open reading frames are found in the first 170 nucleotides of the oriT sequence (Fig. VII.3). Gene X (22), whose function is undefined (3), is transcribed from the opposite strand to that presented here and has a promoter sequence and ribosome binding site (RBS) as noted in Figure 3 corresponding to the sequences proposed by Thompson (22). The initiator codon at position 48 as well as the promoter and RBS sequences are conserved for F, R1 and ColB4. However, the equivalent codon in R100 is changed to ACG possibly preventing translation of gene X. There is another ATG sequence four codons downstream which is in the same reading frame and could possibly initiate gene X translation in R100. The second open reading frame extends from the BglII site to position 167, spanning the oriT nick sites, and could represent a protein which can only be expressed when the two daughter strands have recircularised after conjugal DNA transfer. R100 provides evidence against this possibility since the open reading frame upstream from the BglII site stops unproductively (W. Dempsey, unpublished results). Until further sequence information from F, R1 or ColB4 is obtained, it is unclear whether this open reading frame could be expressed. Cram et al. have shown that several polypeptides are encoded in this region, the best candidates being the 6d protein and a 13.5 kD protein which map very close to oriT (3).

Ostermann et al. have suggested that the inverted repeats at positions 176-206 and 239-274 within and following the AT-rich region in R1 form a recognition site for one of the enzymes involved in conjugal DNA transfer (15). While the position of the first possible hairpin structure is mantained in all four types (Figure VII.3), the patterns of inverted and direct repeats in the next 168 nucleotides (232-400) is unique for each plasmid. In all four types there is a region of dyad symmetry immediately following the A-T rich region which is unique for each plasmid (positions 232-280). This region is a possible candidate for traM protein recognition, as the four different traM products from F, R1, ColB4, and R100 presumably recognize four different but equivalent sequences in the oriT region. However, the four unique traM gene products may recognize any sequence in this region (positions 232-400), as these sequences differ to varying extents for each plasmid.

Thompson et al. (22) reported a sequence in F oriT (positions 236-241) which matched the consensus sequence AAGCGG required for primosome initiation (the n′ protein recognition site [23]). The corresponding sequence in ColB4, R1 and R100 is not conserved nor does it occur in the oriT region presented here. The homology between the oriTs of F and ColE1 reported by Thompson et al. (22) was not conserved in ColB4, R1 or R100. There was also no apparent homology found with oriTs from RSF1010, CloDF13 and RK2 (29).

Comparisons of the predicted traM proteins from ColB4, R1 and R100 to the F traM product (Fig. VII.4, Table VII.1) show that the R100 traM protein is the most homologous to F, followed by ColB4 and then R1. While ColB4 and R1 have been found to be more like each other with respect to traJ, traY and finP, the traM product of ColB4 resembles more closely the traM products from R100 and F than from R1. However, R1 is organized very differently at the DNA level in the region thought to bind the traM product- the 1.1 Kb BglII fragment (29), or, more specifically positions 230 to the start of traM translation. F, ColB4 and R100 are more similar to each other in this region than to R1, and it would be expected that the R1 traM protein would be the least homologous of the four considered here.

While some homologies exist in the putative traM recognition region between F, R100 and ColB4, it should be emphasized that R1, R100 and ColB4 are unable to complement an F traM mutation (Flac traM102 [27]). Thus the few changes seen between these proteins and their putative DNA recognition sequences are sufficient to define four traM alleles (27). Examination of the various traM products revealed no single position that was different for all four known traM proteins (Fig. VII.4). Instead, there was a region that contained the majority (35) of mismatches (residues 1-40), while the other differences (16) were conservative changes distributed throughout the remainder of these proteins (residues 41-127). Thus the amino terminal region (1-40) may dictate traM specificity by recognizing the unique oriT sequence discussed previously.

It has been demonstrated that the F traI product is a DNA helicase I, possessing a DNA-dependent ATPase activity (1). This cytoplasmic product is thought to unwind the plasmid DNA during transfer. The R1 and ColB4 traI products are capable of complementing an F traI mutation, while the F and R100 traI products are not interchangable (26). It has been suggested that the products of traM and traI may form a complex that displaces the traYZ endonuclease (28). If traM and traI proteins interact, one would expect a region (or residue) that is common to the F, ColB4 and R1 traM proteins, but differs in the R100 protein. Although there are several such possibilities, most of these differences are conservative changes. The most striking difference occurs at position 22, where F, ColB4 and R1 contain a positively charged residue, while the R100 product contains the neutral amino acid glutamine.

It has also been suggested that traI and traD may interact (16) since both gene products are thought to provide the motive force necessary to propel the DNA into the recipient cell (1,12). While traM is thought to receive the signal that a competent recipient cell has been engaged and triggers transfer (29), its putative binding site near the end of the retained strand suggests that it may be involved in sensing the initiation and perhaps the conclusion of conjugal DNA metabolism and not the actual physical transfer of DNA.

These sequencing studies on the oriT region of ColB4 complement previous studies on the various alleles for oriT, traM and traY and help illustrate the evolution of transfer genes in conjugative plasmids. While the spatial organization of the transfer genes is approximately the same in all four plasmids, the variation in select parts of the DNA sequence resulting in plasmid-specific alleles is intriguing. These sequencing studies also confirm the heteroduplex mapping studies between F, R1 and R100 (20) which predicited short regions of substitutions in the DNA sequence near oriT.

The evolution of these plasmids' control of DNA metabolism and transfer operon function is fascinating in that the alterations in sequence of the various alleles identified to date (with the possible exception of traJ) are never so great as to obscure function. These studies should be useful in focusing further experimentation towards understanding plasmid replication and transfer during conjugation.

TABLE VII.1

Percent Homologies of oriT/traY/traM Alleles from IncF Plasmids

Plasmids Compared	OriT Nick Region	A-T Rich Region	TraYp Recognition Region	TraMp Recognition Region (?)	traM[a]	TraMp[b]	traY[a]	TraYp[b]
Positions (Fig. 3)	110-167	152-237	174-231	232-391	-	-	-	-
F-ColB4	100	66	53	71	85	86	62	45
F-R1	100	66	50	42	78	78	62	45
F-R100	96	66	57	68	90	89	57	39
ColB4-R1	100	96	96	41	77	76	94	96
ColB4-R100	96	72	57	81	88	88	58	49
R1-R100	96	69	60	46	78	77	60	49

a. DNA sequences of the traM and traY genes were taken from this lab (ColB4 traM, traY; R1 traM, traY; R100 traY) and published results (F traM [21]; R100 traM [5]). The traM gene homology includes positions 464-844 (Fig. 2A) and comparable sequences in F, R1 and R100. The traY gene homology includes positions 77-301 for ColB4 (Fig. 2B), positions 31-255 for R100 (Fig. 2C) and comparable sequences in F and R1. The first 132 nucleotides in F traY were omitted in this homology study since a comparable region in R1, R100 and ColB4 does not exist.

b. The protein homology study of TraMp and TraYp is derived from Fig. 4 and Fig. 5 respectively, where the first 44 amino acid residues of F TraYp are omitted.

FIGURE VII.1

DNA sequencing strategy for _oriT_, _traM_, _finP_, _traY_ and part of _traJ_ from ColB4. Restriction sites for _Aha_III, _Bal_I, _Bgl_II, _Pst_I, _Sau3a_, and _Taq_I used in sequencing studies are marked. The direction and length of sequenced DNA fragments are shown by arrows beneath the restriction map. The relative location of the genes sequenced is also shown.

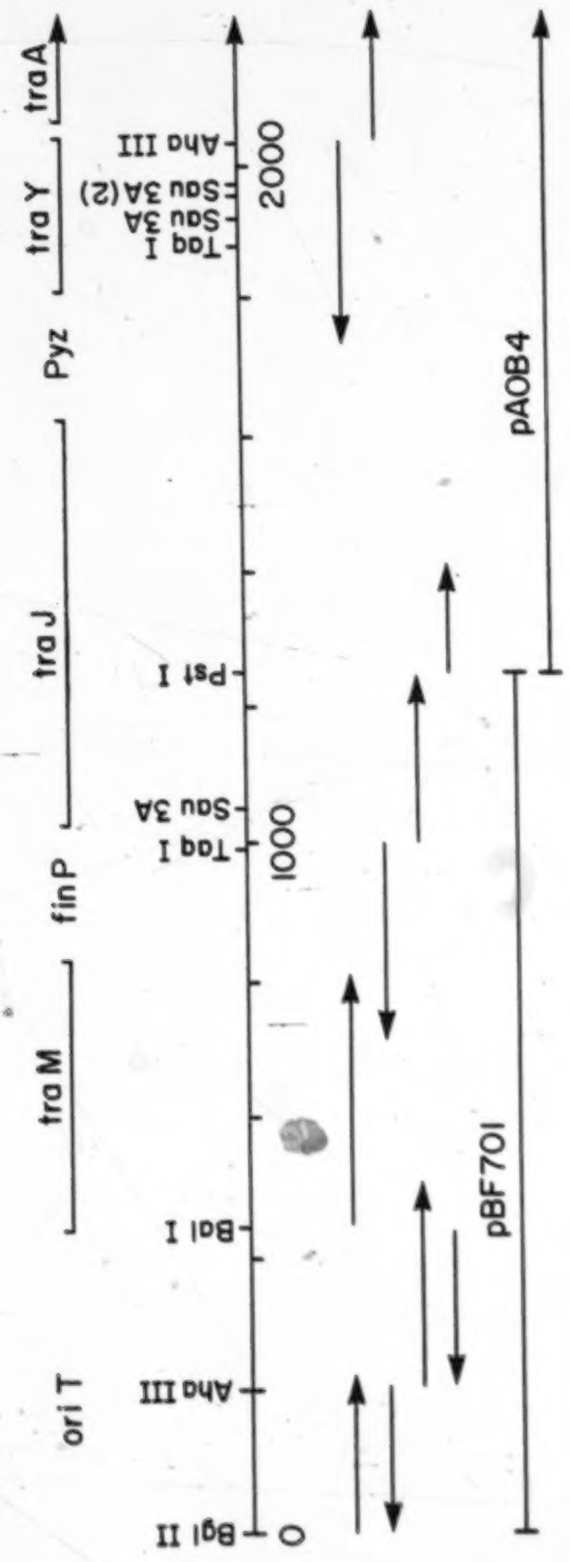

ori T
tra M
fin P
tra J
Pyz
tra Y
tra A
Bgl II
Aha III
Bal I
Taq I
Sau 3A
Pst I
Taq I
Sau 3A
Sau 3A(2)
Aha III
0
1000
2000
pBF701
pAOB4

FIGURE VII.2

a) Nucleotide sequence of oriT and traM from ColB4. Potential -35 and -10 promoter regions and a possible traM ribosome binding site are underlined, and inverted repeats are represented by arrows. The three major oriT nick sites reported by Thompson *et al*. (22) in F and conserved in ColB4 are marked (Λ), although these nicks actually occur in the complementary strand. b) Nucleotide sequence of traY and part of traA from ColB4. The remainder of the traA sequence is published elsewhere (9). Nucleotides which differ in R1 (7) are presented above the ColB4 sequence, and amino acids which differ are located below the translated ColB4 sequence. Potential ribosome sites are underlined. c) Nucleotide sequence of traY from R100-1.

A

```
 BglII                                         50                                                 100
AGATCTCATTTATAAACATCAAGCAGATGGCTAACATCCATTTTTTCATTTTTCCACCTCTGGTGACTTTATCCGTAAATAATTTAACCCACTCCACAAA

                                               150                                                Aha
AAGGCTCAACAGGTTGGTGGTTCTCACCACCAAAAGCACCACACCCCACGCAAAAACAAGTTTTTGCTGATTTGCTTTTTGAATCATTAGCTTATGTTTT
                           ^ ^ ^
III                                            250                                                300
AAATAATGTATTTTAATTTATTTTACATTATTAAATAGGGTCGTTGACGACCCTATGGCGACCCTATGGCGACCCTTGTTAGGGTCAGTACTGACAAATT
                                                                                                -35
                                               350                                                400
TTATGAAAAACATTTGGATATATTAGGGCCGCCCCTAGCAGCGCCGTGCGTTTTTTTATAGTATGACCCTAGGGGCGCTGCTAGGGGTGCTCTTCTTTTT
               -10
                                               450            .TraM BalI.                         500
TCAGACCAAACTATGAGTAAATGAAATAAAACTCTTATTTCAGAATATAAAAGGTACTGTCTTATGGCCAGAGTAAATCTGTATATCAGTAATGAGGTTC
 -35                   -10                                     MetAlaArgValAsnLeuTyrIleSerAsnGluValH
                                                                1                         10

                                               550                                                600
ATGAAAAAATTAACATGATTGTTGAAAAGCGTCGTCAGGAGGGAGCAAGAGATAAAGATATAAGCCTTTCAGGAACTGCTTCAATGCTTCTTGAATTGGG
isGluLysIleAsnMetIleValGluLysArgArgGlnGluGlyAlaArgAspLysAspIleSerLeuSerGlyThrAlaSerMetLeuLeuGluLeuGl
                       20                            30                            40

                                               650                                                700
GCTTCGCGTATATGATGCACAGATGGAGCGTAAAGAGTCTGCGTTTAACCAGACAGAGTTTAATAAACTTCTTCTTGAATGTGTTGTAAAAACACAGTCA
yLeuArgValTyrAspAlaGlnMetGluArgLysGluSerAlaPheAsnGlnThrGluPheAsnLysLeuLeuLeuGluCysValValLysThrGlnSer
             50                            60                            70

                                               750                                                800
ACGGTGGCAAAGATTTTAGGTATTGAGTCTCTCAGTCCTCATGTCTCCGGAAACCCGAAGTTTGAATATGCCAGTATGGTTGACGATATCAGAGAGAAAG
ThrValAlaLysIleLeuGlyIleGluSerLeuSerProHisValSerGlyAsnProLysPheGluTyrAlaSerMetValAspAspIleArgGluLysV
 80                          90                            100                           110

                                               850
TGTCTGTTGAGATGGACCGGTTTTTTCCAAAAAATGATGATGAATAAATGAAAATGAAGTTTGAGTTGGTTCAAATATAAGAATTTTTT
alSerValGluMetAspArgPhePheProLysAsnAspAspGlu***
                       120
```

B

```
         G           T                   G    50                                TraY              100
CCAATTTTGGCTTGTTTTTTTTCGCCGGTTTCTGTGCAATCATAATTTTATTAGTCGGGAAACAGAGGTGAATTGAGTGAGGAGGCGTAACGCGAGAGGC
                                                                           MetArgArgArgAsnAlaArgGly
                                                                            1

 G          A   .A      TaqI                  150    CG.     Sau3A      T.        T    G T
GGTATAAGCAGAACGGTCTCGGTGTATCTCGATGAAGATACAAATAACCGACTTATTAAGGCCAAAGATCGTAGTGGAAGAAGTAAAACGATAGAAGTCC
GlyIleSerArgThrValSerValTyrLeuAspGluAspThrAsnAsnArgLeuIleLysAlaLysAspArgSerGlyArgSerLysThrIleGluValG
  10                         20                          Arg   30                           40

 Sau3A  .   Sau3A.      G.                   T250                 T       A.          AhaIII  300
AGATCCGTCTCAGAGATCATCTGAAACGATTTCCTGATTTCTATAACGAAGAAATATTCAGGGAGGTGACCGAAGAGAGCGAGTCAACATTTAAAGAACT
lnIleArgLeuArgAspHisLeuLysArgPheProAspPheTyrAsnGluGluIlePheArgGluValThrGluGluSerGluSerThrPheLysGluLe
                          50                           60             Ile     Asn   70

                                      TraA A TA T 350                                             400
TTAGTTTTTATCTATCAATATTAAACAGAGGAATCATATGGACGCCGTTTTAAGTGTTCAGGGTGCTTCTGCGCCCGTCAAAAAGAAATCGTTTTTTTCT
u***                                   MetAspAlaValLeuSerValGlnGlyAlaSerAlaProValLysLysLysSerPhePheSer
                                        1 AsnThr                   10                           20
```

C

```
                               TraY                50                                             100
TTTAATTATCTAAGTGGCTGAGGAGGAGACTTGAGCAGAAATATAATAAGGCCTGCTCCGGGAAATAAAGTTTTACTGGTATTGGACGATGCTACTAATC
                               MetSerArgAsnIleIleArgProAlaProGlyAsnLysValLeuLeuValLeuAspAspAlaThrAsnH
                                1                        10                          20

                   SmaI                            150                      Sau3A                 200
ACAAGCTCTTAGGTGCCCGGGAACGTAGTGGGAGAACAAAAACTAATGAGGTGCTAGTCAGGTTACGTGATCATCTGAACCGTTTTCCTGACTTTTATAA
isLysLeuLeuGlyAlaArgGluArgSerGlyArgThrLysThrAsnGluValLeuValArgLeuArgAspHisLeuAsnArgPheProAspPheTyrAs
                     30                          40                           50

                                       TaqI     250Sau3A                                    TraA  300
TCTGGATGCAATAAAGGAGGGAGCAGAGGAAACTGATTCGATAATTAAAGATCTTTAGGTTTATTAATCTATCCACAATAGAGGTAACTTATGAATACTG
nLeuAspAlaIleLysGluGlyAlaGluGluThrAspSerIleIleLysAspLeu***                                  MetAsnThr
            60                          70
```

FIGURE VII.3

Comparison of F (I), ColB4 (II), R1 (III) and R100 (IV) *oriT* sequences. Sequences were aligned by the program of Queen and Korn (17), with spaces inserted to allow maximum alignment. Arrows have been placed over inverted repeats, and double-headed arrows were used when sets of inverted repeats overlap. Open headed arrows (—▷) mark direct repeats while small closed arrows (—▸) mark the short repeats peculiar to ColB4. Potential *traM* promoters are underlined and the n′ primosome recognition site proposed for F (22) is boxed (positions 236-241). The proposed Gene X promoter (which is encoded by the opposite strand), and the ribosome binding sites (RBS) and translational start sites for Gene X and TraMp are boxed. Proposed functional domains are marked, and are discussed in detail in the text. The F and R1 *oriT* sequences were taken from references 22 and 15 respectively, and the R100 sequence was supplied by W. Dempsey (manuscript in preperation).

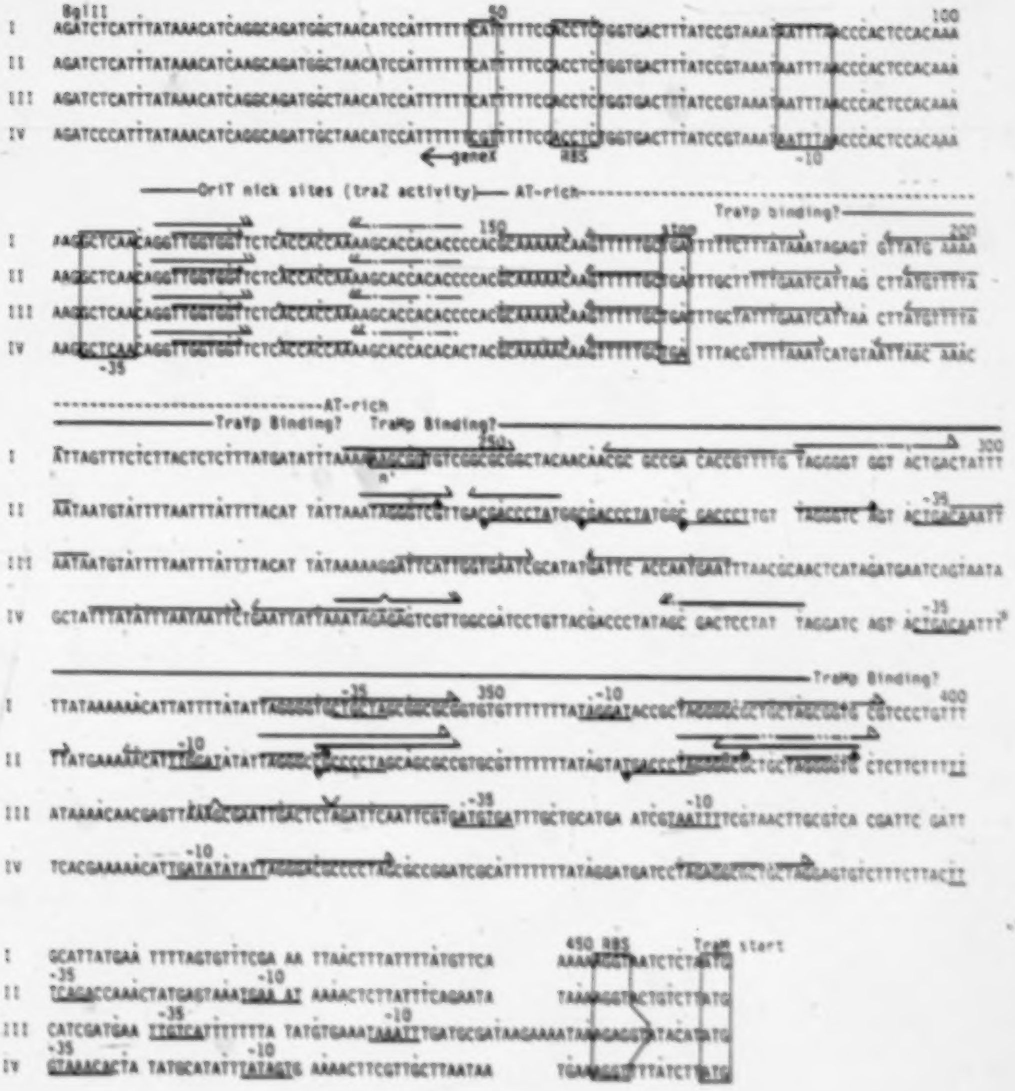

BglII
I
II
III
IV
50
100
gene X
RBS
-10
-35
OriT nick sites (traZ activity)
AT-rich
TraYp binding?
150
200
AT-rich
TraYp Binding?
TraMp Binding?
250
300
TraMp Binding?
350
400
450
RBS
TraM start

FIGURE VII.4

Comparison of F, R1, ColB4 and R100 traM open reading frames. Non-homologous sequences are boxed. TraM sequences are from references 21,13,7,5 respectively.

										10										20
F	Met	Ala	Lys	Val	Asn	Leu	Tyr	Ile	Ser	Asn	Asp	Ala	Tyr	Glu	Lys	Ile	Asn	Ala	Ile	Ile
R1	Met	Ala	Lys	Val	Gln	Ala	Tyr	Val	Ser	Asp	Glu	Ile	Val	Tyr	Lys	Ile	Asn	Lys	Ile	Val
ColB4	Met	Ala	Arg	Val	Asn	Leu	Tyr	Ile	Ser	Asn	Glu	Val	His	Glu	Lys	Ile	Asn	Met	Ile	Val
R100	Met	Ala	Arg	Val	Ile	Leu	Tyr	Ile	Ser	Asn	Asp	Val	Tyr	Asp	Lys	Val	Asn	Ala	Ile	Val

										30										40
F	Glu	Lys	Arg	Arg	Gln	Glu	Gly	Ala	Arg	Glu	Lys	Asp	Val	Ser	Phe	Ser	Ala	Thr	Ala	Ser
R1	Glu	Arg	Arg	Arg	Ala	Glu	Gly	Ala	Lys	Ser	Thr	Asp	Val	Ser	Phe	Ser	Ser	Ile	Ser	Thr
ColB4	Glu	Lys	Arg	Arg	Gln	Glu	Gly	Ala	Arg	Asp	Lys	Asp	Ile	Ser	Leu	Ser	Gly	Thr	Ala	Ser
R100	Glu	Gln	Arg	Arg	Gln	Glu	Gly	Ala	Arg	Asp	Lys	Asp	Ile	Ser	Val	Ser	Gly	Thr	Ala	Ser

										50										60
F	Met	Leu	Leu	Glu	Leu	Gly	Leu	Arg	Val	His	Glu	Ala	Gln	Met	Glu	Arg	Lys	Glu	Ser	Ala
R1	Met	Leu	Leu	Glu	Leu	Gly	Leu	Arg	Val	Tyr	Glu	Ala	Gln	Met	Glu	Arg	Lys	Glu	Ser	Ala
ColB4	Met	Leu	Leu	Glu	Leu	Gly	Leu	Arg	Val	Tyr	Asp	Ala	Gln	Met	Glu	Arg	Lys	Glu	Ser	Ala
R100	Met	Leu	Leu	Glu	Leu	Gly	Leu	Arg	Val	Tyr	Glu	Ala	Gln	Met	Glu	Arg	Lys	Glu	Ser	Ala

										70										80
F	Phe	Asn	Gln	Thr	Glu	Phe	Asn	Lys	Leu	Leu	Leu	Glu	Cys	Val	Val	Lys	Thr	Gln	Ser	Ser
R1	Phe	Asn	Gln	Ala	Glu	Phe	Asn	Lys	Val	Leu	Leu	Glu	Cys	Ala	Val	Lys	Thr	Gln	Ser	Thr
ColB4	Phe	Asn	Gln	Thr	Glu	Phe	Asn	Lys	Leu	Leu	Leu	Glu	Cys	Val	Val	Lys	Thr	Gln	Ser	Thr
R100	Phe	Asn	Gln	Thr	Glu	Phe	Asn	Lys	Leu	Leu	Leu	Glu	Cys	Val	Val	Lys	Thr	Gln	Ser	Ser

										90										100
F	Val	Ala	Lys	Ile	Leu	Gly	Ile	Glu	Ser	Leu	Ser	Pro	His	Val	Ser	Gly	Asn	Ser	Lys	Phe
R1	Val	Ala	Lys	Ile	Leu	Gly	Ile	Glu	Ser	Leu	Ser	Pro	His	Val	Ser	Gly	Asn	Pro	Lys	Phe
ColB4	Val	Ala	Lys	Ile	Leu	Gly	Ile	Glu	Ser	Leu	Ser	Pro	His	Val	Ser	Gly	Asn	Pro	Lys	Phe
R100	Val	Ala	Lys	Ile	Leu	Gly	Ile	Glu	Ser	Leu	Ser	Pro	His	Val	Ser	Gly	Asn	Pro	Lys	Phe

										110										120
F	Glu	Tyr	Ala	Asn	Met	Val	Glu	Asp	Ile	Arg	Glu	Lys	Val	Ser	Ser	Glu	Met	Glu	Arg	Phe
R1	Glu	Tyr	Ala	Asn	Met	Val	Glu	Asp	Ile	Arg	Asp	Lys	Val	Ser	Ser	Glu	Met	Glu	Arg	Phe
ColB4	Glu	Tyr	Ala	Ser	Met	Val	Asp	Asp	Ile	Arg	Glu	Lys	Val	Ser	Val	Glu	Met	Asp	Arg	Phe
R100	Glu	Tyr	Ala	Asn	Met	Val	Glu	Asp	Ile	Arg	Glu	Lys	Val	Ser	Ser	Glu	Met	Glu	Arg	Phe

							127
F	Phe	Pro	Lys	Asn	Asp	Asp	Glu
R1	Phe	Pro	Glu	Asn	Asp	Glu	Glu
ColB4	Phe	Pro	Lys	Asn	Asp	Asp	Glu
R100	Phe	Pro	Lys	Asn	Asp	Glu	Glu

FIGURE VII.5

Comparison of F, R1, ColB4 and R100 _traY_ predicted gene products. Non-homologous sequences are boxed. The F and R1 _traY_ sequences are from references 8 and 7 respectively.

```
                                            10                                      20
F      Met Val Lys Leu Lys Leu Pro Val Asp Val Glu Ser Leu Leu Ile Glu Ala Ser Asn Arg

                                            30                                      40
F      Ser Gly Arg Ser Arg Ser Phe Glu Ala Val Ile Arg Leu Lys Asp His Leu His Arg Tyr

                                            50                                      60
F      Pro Lys Phe Asn Arg Ala Gly Asn Ile Tyr Gly Lys Ser Leu Val Lys Tyr Leu Thr Met
ColB4                  Met Arg Arg Arg Asn Ala Arg Gly Gly Ile Ser Arg Thr Val Ser Val
R1-19                  Met Arg Arg Arg Asn Ala Arg Gly Gly Ile Ser Arg Thr Val Ser Val
R100                   Met Ser Arg Asn Ile Ile Arg Pro Ala Pro Gly Asn Lys Val Leu Leu
                                                            10

                                            70                                      80
F      Arg Leu Asp Asp Glu Thr Asn Gln Leu Leu Ile Ala Ala Lys Asn Arg Ser Gly Trp Cys
ColB4  Tyr Leu Asp Glu Asp Thr Asn Asn Arg Leu Ile Lys Ala Lys Asp Arg Ser Gly Arg Ser
R1-19  Tyr Leu Asp Glu Asp Thr Asn Asn Arg Leu Ile Arg Ala Lys Asp Arg Ser Gly Arg Ser
R100   Val Leu Asp Asp Ala Thr Asn His Lys Leu Leu Gly Ala Arg Glu Arg Ser Gly Arg Thr
                    20                                      30

                                            90                                     100
F      Lys Thr Asp Glu Ala Ala Asp Arg Val Ile Asp His Leu Ile Lys Phe Pro Asp Phe Tyr
ColB4  Lys Thr Ile Glu Val Gln Ile Arg Leu Arg Asp His Leu Lys Arg Phe Pro Asp Phe Tyr
R1-19  Lys Thr Ile Glu Val Gln Ile Arg Leu Arg Asp His Leu Lys Arg Phe Pro Asp Phe Tyr
R100   Lys Thr Asn Glu Val Leu Val Arg Leu Arg Asp His Leu Asn Arg Phe Pro Asp Phe Tyr
                    40                                      50

                                           110                                     119
F      Asn Ser Glu Ile Phe Arg Glu Ala Asp Lys Glu Glu Asp Ile     Thr Phe Asn Thr Leu
ColB4  Asn Glu Glu Ile Phe Arg Glu Val Thr     Glu Glu Ser Glu Ser Thr Phe Lys Glu Leu
R1-19  Asn Glu Glu Ile Phe Arg Glu Val Ile     Glu Glu Asn Glu Ser Thr Phe Lys Glu Leu
R100   Asn Leu Asp Ala Ile Lys Glu Gly Ala     Glu Glu Thr Asp Ser Ile Ile Lys Asp Leu
                    60                                          70
```

E. BIBLIOGRAPHY

1. **Abdel-Monem, M., G. Taucher-Scholz, and M.-Q. Klinkert**. 1983. Identification of Escherichia coli DNA helicase I as the traI gene product of the F sex factor. Proc. Natl. Acad. Sci. U.S.A. **80**:4659-4663.

2. **Birnboim, H.C., and J. Doly**. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1523.

3. **Cram, D., A. Ray, L. O'Gorman, and R. Skurray**. 1984. Transcriptional analysis of the leading region in F plasmid DNA tranfer. Plasmid. 11:221-233.

4. **Everett, R., and N. Willetts**. 1980. Characterization of an in vivo system for nicking at the origin of conjugal DNA transfer of the sex factor F. J. Mol. Biol. **136**:129-150.

5. **Fee, B.E., and W.B. Dempsey**. The cloning, mapping and sequencing of the plasmid R100 traM amd finP genes. J. Bacteriol. submitted.

6. **Finlay, B.B., L.S. Frost, and W. Paranchych**. 1984. Localization, cloning and sequence determination of the conjugative plasmid ColB2 pilin gene. J. Bacteriol. **160**:402-407.

7. **Finlay, B.B., L.S. Frost, and W. Paranchych**. 1986. Nucleotide sequence of the R1-19 plasmid transfer genes traM, finP, traJ, traY, and traYZ promoter. J. Bacteriol. May issue.

8. **Fowler, T., L. Taylor, and R. Thompson**. 1983. The control region of the F plasmid transfer operon: DNA sequence of the traJ and traY genes and characterization of the traYZ promoter. Gene **26**:79-89.

9. **Frost, L.S., B.B. Finlay, A. Opgenorth, W. Paranchych and J. Lee**. 1985. Characterization and sequence analysis of pilin from IncF plasmids. J. Bacteriol. **164**:1238-1247.

10. **Hawley, D., and W. McClure**. 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res. **11**:2237-2255.

11. **Humphries, G., G.A. Willshaw, and E.S. Anderson**. 1975. A simple method for the preparation of large quantities of pure plasmid DNA. Biochim. Biophys. Acta **383**:457-463.

12. **Kingsman, A., and N. Willetts**. 1978. The requirements for conjugal DNA synthesis in the donor strain during Flac transfer. J. Mol. Biol. **122**:287-300.

13. **Koronakis, V., E. Bauer, and G. Hogenauer**. 1985. The traM gene of the resistance plasmid R1: comparison with the corresponding sequence of the Escherichia coli F factor. Gene **36**:79-86.

14. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1-45.

15. **Ostermann, E., F. Kricek, and G. Hogenauer**. 1984. Cloning the origin of tranfer region of the resistance

plasmid R1. EMBO J. 3:1731-1735.

16. **Panicker, M.M., and E.G. Minkley, Jr.** 1985. DNA transfer occurs during a cell surface contact stage of F sex factor-mediated bacterial conjugation. J. Bacteriol. 162:584-590.

17. **Queen, C., and L.J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids. Res. 12:581-599.

18. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13:319-353.

19. **Sanger, F., S. Nicklen, and A.R. Coulson.** 1977. DNA sequencing with chain terminating-inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

20. **Sharp, P.A., S.N. Cohen, and N. Davidson.** 1973. Electron microscope heteroduplex studies of sequence relations among plasmids of _E_. _coli_ II. Studies of drug resistant (R) factors and F factors. J. Mol. Biol. 75:235-255.

21. **Thompson, R., and L. Taylor.** 1982. Promoter mapping and DNA sequencing of the F plasmid transfer genes _traM_ and _traJ_. Mol. Gen. Genet. 188:513-518.

22. **Thompson, R., L. Taylor, K. Kelly, R. Everett, and N. Willetts.** 1984. The F plasmid origin of transfer: DNA sequence of wild-type and mutant origins and location of origin-specific

nicks. EMBO J. 3:1175-1180.

23. **Van der Ende, A., R. Teerstra, H.G.A.M. Van der Avoort, and P.J. Weisbeek**. 1983. Initiation signals for complementary strand DNA synthesis on single-stranded DNA. Nucleic Acids Res. 11:4957-4975.

24. **Vieira, J., and J. Messing**. 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259-268.

25. **Wallace, R.B., M.J. Johnson, S.V. Suggs, K. Miyoshi, R. Bhatt, and K. Itakura**. 1981. A set of synthetic oligodeoxyribonucleotide primers for DNA sequencing in the plasmids vector pBR322. Gene **16**:21-26.

26. **Willetts, N., and J. Maule**. 1979. Investigations of the F conjugation gene traI: traI mutants and lambda traI transducing phages. Mol. Gen. Genet. **169**:325-336.

27. **Willetts, N.S., and J. Maule**. 1986. Specificities of incF plasmid conjugation genes. Genet. Res. **47**:1-11.

28. **Willetts, N.S., and R. Skurray**. 1980. The conjugation system of F-like plasmids. Ann. Rev. Genet. **14**:41-76.

29. **Willetts, N., and B. Wilkins**. 1984. Processing of plasmid DNA during bacterial conjugation. Microbiol. Rev. **48**:24-41.

Chapter VIII

NUCLEOTIDE SEQUENCE OF THE SURFACE EXCLUSION GENES traS AND traT FROM THE IncFV PLASMID pED208[1]

A. INTRODUCTION

Surface (or entry) exclusion (sfx) can be defined as the activity that reduces the ability of plasmid containing cells to participate as recipients in conjugal matings with other cells containing a closely related or identical plasmid (19). This property is unrelated to plasmid incompatibility which prevents the co-existance of two plasmids within a cell (28). The F plasmid encodes two surface exclusion genes, traS and traT, which provide exclusion activity by two independent mechanisms (1). These genes are located within the major F transfer (traYZ) operon between traG and traD, immediately following the region required for F pilus biosynthesis (21). The traS gene from F encodes an 18,000 dalton inner membrane protein (1) which may affect triggering of conjugal DNA metabolism (20) since traS is capable of inhibiting DNA transfer even

[1] A version of this chapter has been accepted. **Finlay, B.B., and W. Paranchych**. 1986. J. Bacteriol. June issue.

when stable mating aggregates form. The product of traT, TraTp, is a 25,000 dalton protein found in large amounts (greater than 20,000 copies per cell) in the outer membrane of F plasmid containing cells (1). This protein provides surface excusion by reducing stable mating aggregate formation (1). It has recently been shown that TraTp from F is a lipoprotein (25).

The nucleotide sequence of traT from R-100 has been published (24) and the predicted TraTp is highly homologous to TraTp from F (22). It has been demonstrated that the traT gene products from R6-5 (whose conjugal system is closely related to R-100) and R-100 mediate serum resistance (23) and, at least in the case of R6-5, TraTp enhances resistance of E. coli to phagocytosis (4).

pED208 (formerly EDP208) is a derepressed derivative of a naturally occurring lac plasmid, Folac, originally identified in Salmonella typhi (8). F and pED208 pili share no serological similarity and it has been suggested that pED208 does not contain a F-like transfer operon (33). Cells carrying pED208 or F-like plasmids are sensitive to the pilus-specific phage f1 while only cells containing F-like plasmids are sensitive to the RNA phages f2 and R17 (5,6). In this chapter I describe cloning and sequencing of the surface exclusion system from the pED208 plasmid.

B. MATERIALS AND METHODS

1. Bacterial strains and plasmids.

Plasmids pED208 and F were carried in the host strain JC6256 (E. coli K-12 F⁻ trp lac) and kindly donated by N. S. Willetts (Biotechnology Australia, N.S.W., Australia). E. coli P67854 (3) was the strain used for minicells experiments. Recombinants were transformed into JM83 (31) or AA125 (his trp bio Str^r) and JM103 was used for M13 cloning and sequencing (35). pED851 is a chimera consisting of the large BamHI fragment from F (containing the entire F tra operon) cloned into pBR322 (17). Chimeras constructed are described in Table VIII.I.

2. Recombinant DNA techniques and DNA sequencing.

These were performed as described previously (10,13).

3. Isolation of minicells.

[^{35}S]-methionine labelled minicells were prepared as described elsewhere (2) followed by electrophoresis and autoradiography.

4. Surface exclusion assay and index.

Standing overnight cultures of donor and recipients were diluted twenty fold in L-broth, grown to mid-logarithmic phase (5×10^8 cells/ml) and stored on ice until required. One ml of donor (*lac*$^+$) was added to an equivalent amount of recipient (*lac*$^-$) and allowed to sit for 1 hr, 37°. Various dilutions were plated onto lactose base MacConkey agar containing appropriate antibiotic (which prevents growth of donor cells) and lac$^+$ and lac$^-$ colonies were counted. Approximately 1000 colonies (*lac*$^-$ and *lac*$^-$) were counted from each mating mixture. Concurrent matings with a surface exclusion negative recipient (pBF106/JM83) and with test strains were used to measure the surface exclusion index. The surface exclusion index was calculated by the formula:

$$\text{sfx index} = \frac{\dfrac{\text{Transconjugants}_{\text{pBF106}}}{\text{Transconjugants}_{\text{pBF106}} + \text{Recipients}_{\text{pBF106}}}}{\dfrac{\text{Transconjugants}_{\text{test}}}{\text{Transconjugants}_{\text{test}} + \text{Recipients}_{\text{test}}}}$$

where "Transconjugants$_{\text{pBF106}}$" is the number of colonies containing both pBF106 [surface exclusion negative; (11)] and the donor (either pED208 (*lac*$^+$) or F*lac*) plasmids, "Recipients$_{\text{pBF106}}$" is the number of colonies containing only pBF106 (*lac*$^-$), "Transconjugants$_{\text{test}}$" is the number of colonies containing both the chimera being tested and the donor plasmid, and "Recipients$_{test}$" is the number of colonies containing only the test chimera. A value of one indicates no surface exclusion activity, while larger numbers indicate the presence of the surface exclusion function.

C. RESULTS

1. Cloning of the pED208 surface exclusion system.

We have previously described cloning of the four large HindIII fragments from the pED208 plasmid into pACYC184 [pBF101-pBF104;(11)], with their locations illustrated in Fig. 1. When these four fragments (which contain 93% of the pED208 plasmid DNA) were tested for surface exclusion activity towards pED208, only pBF101 exhibited detectable activity. pBF101 is the largest of these HindIII clones and encodes the genes required for pilus biosynthesis (11), and presumably this surface exclusion activity is part of the pED208 transfer operon, as it is in F. Subcloning of pBF101 localized surface exclusion activity to a 7.5 Kilobase (Kb) XhoI fragment (pBF111) and then to a 3.7 Kb SalI/XhoI fragment (pBF120), which is adjacent to the area required for pilus production (Fig. VIII.1 and Table VIII.2).

The SalI/XhoI fragment from pBF120 was extensively subcloned as shown in Fig. VIII.2. Of all the recombinants constructed, only chimeras containing three restriction fragments demonstrated surface exclusion (Table VIII.2, Fig. VIII.2). pBF126 (which contains a 1.3 kb PstI fragment inserted into pBR322) exhibited exclusion of pED208, while the same fragment ligated in the opposite orientation (pBF125) did not exclude pED208. When this PstI fragment was introduced into pUC8, chimeras containing one orientation excluded pED208 (pBF137), while chimeras with the opposite orientation (pBF136) were exclusion negative. From this information the direction of transcription of this surface exclusion area can be determined. However, the SalI/XhoI 3.7 kb fragment was not placed in the same orientation as a vector promoter in pBF120 but still gave surface exclusion, suggesting a surface exclusion promoter is present in pBF120. When this SalI/XhoI fragment was ligated into pUC8, again a recombinant not requiring a vector promoter excluded pED208 (pBF134). When this insert was placed in the opposite orientation under the lac promoter (pBF135), an exclusion positive phenotype was observed, but the cells were extremely unhealthy, and an accurate exclusion index was not obtained. The other two subclones of pBF120 that expressed surface exclusion were pBF140 and pBF141. Both these chimeras contain the 1.5 kb AhaIII fragment in opposite orientations in pUC vectors. Since pBF140 was not placed under a vector promoter, the exclusion activity encoded by this chimera is expressed independently of a vector promoter and presumably contains its own promoter.

2. DNA Sequencing Studies.

To define the components of the pED208 surface exclusion system, the complete nucleotide sequence of the 3,725 nucleotide _SalI_/_XhoI_ restriction fragment from pBF120 was determined using the M13 dideoxy sequencing method (27) and is shown in Fig. VIII.3. Since the _PstI_ 1.3 kb fragment encodes surface exclusion activity when expressed by a vector promoter (pBF126, pBF137), but does not do so when oriented in the opposite direction (pBF125, pBF136), the direction of transcription should proceed from the _SalI_ site towards the _XhoI_ site (rightward in Fig. VIII.2). If the genes of the transfer operon of pED208 are organized in a similar manner to F, the surface exclusion genes should be found downstream of the area required for pilus biosynthesis (34). As seen in Fig. VIII.1, this is precisely the location of the pED208 exclusion loci, suggesting that the direction of transcription and translation proceeds from the _SalI_ site towards the _XhoI_ site.

When the sequence from Fig. VIII.3 is translated in this direction, four major open reading frames (ORFs) are found. Beginning at the _SalI_ site, there is an open reading frame of 414 amino acids which stops at nucleotide

1244 (Fig. VIII.3). Following this is an 193 residue ORF (nucleotides 1297-1875) which contains the gene we have called traS (for reasons to be explained later). Following traS is an 179 bp intercistronic region and the gene traT, which encodes a 245 amino acid protein (nucleotides 2055-2789). Beginning at position 2883 and proceeding to the XhoI site is another ORF, which, for reasons discussed later, appears to encode 244 amino acids up to and including the XhoI site. There were no major ORFs found when this fragment was translated in the opposite direction.

3. The 414 residue ORF.

Several attempts were made to visualize a protein encoded by the first 1300 nucleotides, but cloning of this region appeared to be deleterious to cells. Although it was possible to clone the SalI/AhaIII 1.4 kb fragment (positions 1-1366) into pUC8 (pBF139), bacteria harboring this recombinant were very slow in growth. Attempts to reverse the orientation of this insert such that it would be placed under the vector lac promoter were unsuccessful, although we were able to place it into pBR328, under the chloramphenicol promoter (pBF148). When we examined minicells containing pBF139 and pBF148 there were no apparent additional proteins produced (data not shown). When the SmaI/AhaIII 1.1 kb fragment (nucleotides 244-1366) was inserted into pUC8 (pBF138), the large ORF is connected in phase to the beginning of the lacZ gene, creating a fusion polypeptide of 338 amino acids. Again, cells containing this chimera were slow in growth. Minicells containing pBF138 produced a polypeptide of 37,000 daltons. Placing this insert into pBR328 under the chloramphenicol promoter removed this protein, suggesting that the polypeptide seen with pBF138 is a fusion protein.

It is not clear why cells harboring recombinants containing this region grow poorly, even when there is no vector promoter transcribing this region (for example, pBF139). There is a possible translational start site a nucleotide 306, preceeded by two possible ribosome binding sites [AGAG and CAGGA; (29)] beginning at nucleotides 295 and 300 repectively. A translational start at this position would create a 312 residue polypeptide, assuming there is an active promoter upstream. As mentioned above, no polypeptides of this size were visible in minicells containing pBF139.

If the tra operons of pED208 and F are ordered similarily, traG should be located immediately preceeding traS. Since traG is large [encoding a 100-116 Kdalton protein; (32)] it is possible that the 414 amino acid ORF is the carboxy-terminal region of an pED208 equivalent of traG.

4. _TraS_.

Bacteria containing the recombinants pBF126 or pBF137 both exhibited surface exclusion activity (see Table VIII.2). Examining the nucleotide sequence spanning this _PstI_ fragment (nucleotides 966-2262) reveals only one major ORF encompassing nucleotides 1297-1875. A probable translational start site for this ORF is the ATG located at position 1318. Finishing 5 nucleotides upstream of this methionine is a sequence (CAGGAGG) which is highly homologous to the consensus ribosome binding site [AGGAGGT; (29)]. Translation of a protein beginning with this methionine would yield a 186 residue polypeptide (21,230 daltons; Fig. VIII.3). No polypeptide corresponding to this ORF was detectable in minicells containing pBF125, pBF126, pBF136, and pBF137.

This predicted protein is highly hydrophobic, with 17% of its residues being charged and 50% being aromatic residues and isoleucine, leucine, methionine or valine. When secondary structure predictions of this protein are examined (Fig. VIII.4), these predictions resemble those found for an integral membrane protein. This protein contains stretches of 20-25 amino acids which are hydrophobic, with small intervening hydrophilic regions. Furthermore, these hydrophilic regions are all predicted to be in the beta-turn conformation (Fig. VIII.4), while the hydrophobic sequences are predicted to be either in beta-sheet or alpha-helix conformations.

The traS product from F is reported to be an 18,000 dalton inner membrane protein produced in small quantities (1). Since this 186 residue polypeptide is similar in size and probably in cellular location, we have called this protein the traS gene product from pED208. As in the F system, the pED208 traS immediately preceeds traT (see later) and follows the region required for pilus biosynthesis (11).

Since bacteria harboring chimeras containing the PstI 1.3 kb fragment in opposite orientation to the vector promoter (pBF125 and pBF136) were unable to exclude pED208 (Table VIII.2), this suggests that there is no promoter upstream of traS, at least up to the PstI site at position 966.

5. TraT.

When the AhaIII 1.5 kb fragment (positions 13 6-2824) was cloned in both orientations into the pUC vectors (pBF140 and pBF141), bacteria containing these chimeras were capable of excluding pED208 (Table VIII.2). Since this AhaIII fragment does not encode the entire traS gene, these observations suggest that pBF140/pBF141 contain another surface exclusion gene, with its own promoter. Minicells of pBF120, pBF140 and pBF141 all produced a 25,000 dalton protein in large quantities (Fig. VIII.5).

When minicells were fractionated, this 25,000 dalton protein was found mainly in the membrane fraction (data not shown). The traT gene products of F and R-100 plasmids encode a 25,000 dalton protein which is located in the outer membrane in high copy number (21), suggesting that this protein from pED208 may be a TraTp equivalent.

Upon examination of the nucleotide sequence of this AhaIII fragment, a large ORF was found (in addition to the truncated traS reading frame) encompassing nucleotides 2001 to 2789. When this translated sequence is compared with the traT gene product from R-100 (24), a highly homologous protein from pED208 (which we have called TraTp) can be aligned with TraTp from R-100 (Fig. VIII.6). The predicted gene product of traT from pED208 begins at nucleotide 2055 and proceeds to nucleotide 2789 (25,981 daltons; Fig. VIII.3). Preceeding the putative start of this protein is a weak ribosome binding site (GAGA) eight nucleotides upstream.

The homology of traT gene products from pED208, R-100, and F is striking (Fig. VIII.6). TraTps from pED208 and R-100 are 80% homologous, with 46 mismatches and 2 unmatched amino acids. The proposed signal sequences differ significantly, with 7 mismatches and 2 unmatched residues out of 22 amino acids. However, the proposed signal cleavage site (Gly_{22}/Cys_{23}; E. Minkley, personal communication) is conserved as is the location of glycerol and fatty acid modification in TraTp (Cys_{23}; (25) and E. Minkley,personal communication). The remaining differences are interspersed thoughout the proteins.

The production of high levels of TraTp independent of vector promoter (i.e. pBF120 and pBF140) suggests that traT encodes its own promoter. Contained within the 179 nucleotides between traS and traT is a possible promoter centered around position 1950 (Fig. VIII.3) which resembles the consensus procaryotic promoter (15). The -35 region contains four of the six consensus sequence nucleotides (TTGATG) as does the -10 region (TACATT). There are 18 nucleotides found between the -10 and -35 regions.

6. The 244 amino acid ORF.

Beyond traT, beginning at nucleotide 2883 is an ORF which continues into the XhoI site at position 3720. There is a potential translational start at nucleotide 2991, preceeded by a ribosome binding sequence (CAGG) eight nucleotides upstream. A translational start at this position would produce an ORF of at least 244 amino acids within the pED208 plasmid. Attempts to detect the production of a polypeptide of this size in minicells using pBF142 were unsuccessful (data not shown). In the F plasmid, traD follows traT, producing a 77-85 Kdalton protein (18). Thus, if the genetic organization of the pED208 tra operon is similar to that of F, this 244 residue

ORF may be the amino-terminal region of the pED208 traD gene product.

7. Cross-reactivity between pED208 and F surface exclusion systems.

Besides the homology found between TraTps from F and pED208, these plasmids encode conjugative pili that are similar in size, contain an acetylated amino terminus (12,14), and can be infected by the pilus specific f1 phage (5). Since surface exclusion may involve, at least partly, blockage of pilus attatchment (1,13), we decided to test for any cross-reactivity between surface exclusion systems from F and pED208. To determine this we utilized recombinants of pED208 and F that encode part, or all of their tra operons. The results of these assays are shown in Table VIII.2.

When recombinants containing the pED208 surface exclusion system (pBF134, pBF137) were tested for exclusion of the pED208 plasmid, high levels of surface exclusion were seen . However, when the recipient contained chimeras encoding the entire F tra operon [pED851; (17)], or the F surface exclusion loci [pRS31; (30)] levels of exclusion only slightly greater than background were obtained. When the F lac plasmid was used as the donor, recipient cells containing the F surface exclusion genes (pED851, pRS31) produced high surface exclusion indices, but low levels of exclusion were seen with recipient cells containing recombinants of the pED208 surface exclusion system (pBF134, pBF137). These results suggest that there may be a small amount of surface exclusion cross-reactivity between F and pED208, but it is significantly smaller than the homologous exclusion systems, and may be insignificant *in vivo*.

D. DISCUSSION

The regulation of *traT* in F-like transfer systems is not completely understood. The TraTp of R-100 has been reported to be produced in large quantities regardless of the transcriptional state of the *tra* operon [i.e. repressed or derepressed (9)]. By sequencing *traT* from R-100, Ogata *et al*. found a promoter sequence upstream of R-100 TraTp. Rashtchian *et al*. (26) reported that TraTp production in F-plasmid containing cells is independent of *traJ* product, which is a positive regulator of the *traYZ* operon, although recent sequence information from the F plasmid suggests that there is no promoter for *traT* between *traS* and *traT* (E. G. Minkley, Jr., personal communication). The high levels of production of TraTp from clones containing pED208 *traT*, in the absence of a vector promoter, suggest that the pED208 system contains a *traT* promoter, analogous to the R-100 system. Both systems contain two nucleotides in the

-35 region of their putative promoters which do not match the procaryotic promoter consensus sequence. R-100 has only one mismatch in the -10 region when compared to the consensus sequence, while pED208 has two. TraS from pED208 does not appear to have a promoter immediately upstream, and presumably it is under traYZ operon control.

The ribosome binding site of traT (GAGA) demonstrates poor homology with the consensus Shine-Dalgarno sequence [AGGAGGT; (29)]. The synthesis of high levels of TraTp is probably due to a strong promoter which produces large amounts of traT mRNA. In contrast, traS has a ribosome binding site which is highly homologous to the consensus sequence, but the traS gene product was not detectable in a mini-cell system. The role of traS in surface exclusion is believed to be involved in the inhibition of DNA transfer (20).

Placement of the surface exclusion genes under different vector promoters greatly varied exclusion capabilities (Table VIII.2). The difference in exclusion levels between pBF126 and pBF137 may be attributed to different vector promoter strengths. TraS in pBF126 is promoted by the ampicillin promoter in pBR322 while in pBF137 it is transcribed by the lac promoter. Placement of traT under the lac promoter (pBF141) resulted in slow growing bacteria, probably due to a higher expression of TraTp. The highest level of surface exclusion involving recombinants of the pED208 surface exclusion system was produced by pBF134. This chimera contains both traS and traT in a pUC8 vector, but they are not transcribed by the vector promoter. When this restriction fragment was placed in the opposite orientation (pBF135), it provided exclusion, but was too unhealthy to obtain an exclusion index. This lethality may be due to high levels of the gene products of traS and traT, or it may also be due to the ORF found upstream of traS, which appears to be fatal when placed under a vector promoter (see Results section).

Minkley and Willetts reported the production of a 57,000 dalton polypeptide when only the carboxy-terminus fragment of traG from F was present (22). These workers attributed it to be a fusion product with the vector or a translational restart product. Although no polypeptides appeared to be produced by the equivalent fragment from pED208 (pBF139, pBF148), a translational restart within the 414 amino acid ORF might explain why this region of DNA is difficult to clone, and fatal if transcribed with the aid of a strong vector promoter, such as the lac promoter (pBF148 was transcribed by the chloramphenicol promoter in pBR328). In F, traG appears to have several functions, including pilus biosynthesis, stabilization of mating pairs and DNA transfer (32). If this gene is indeed the traG equivalent in pED208, a translational restart within this gene may define two functional domains of this protein.

The homology between TraTps from R-100 and pED208 was unexpected, since pED208 was thought to contain a completely unrelated transfer operon. The fact that they have conserved TraTps and that the order and location of the surface exclusion loci is maintained suggests a common evolutionary origin. The difference between processed TraTps from F and R-100 is one amino acid (Fig. 5), yet this difference is sufficient to provide plasmid specificity of these *traT* gene products. The numerous changes in the pED208 TraTp sequence could easily account for a new *traT* specificity as seen by the lack of exclusion cross-reactivity between F and pED208 (Table VIII.2). However, with the large homologies in TraTp sequences from pED208 and F, these conjugative plasmids are more related than first believed.

It has been suggested that a possible receptor for TraTp is the pilus of a donor cell (13). In the IncF plasmid systems it has been shown that pilus type matches *traT* alleles (13,33). The differences in pilus type are due mainly to changes occurring at the carboxy- and amino-termini of processed pilin (13). Since the amino terminal region of pED208 pilin is completely different from that of F-like pilins (12), an altered *traT* sequence was expected. As more *traT* sequences are accumulated, the differences and similarities should lead to a better understanding of this important conjugal protein.

Table VIII.1

Chimeras constructed from the pED208 plasmid

Chimera (reference)	Vector/ Restriction Sites	Fragment inserted (kb)	Figure illustrated in
pBF101 (11)	pACYC184/*Hind*III	*Hind*III 30	1
pBF106 (11)	pBR322/*Hind*III/*Sal*I	*Hind*III/*Xho*I 14	1
pBF111 (11)	pBR322/*Sal*I	*Xho*I 7.51	1
pBF120	pBR322/*Sal*I/*Bam*HI	*Sal*I/*Xho*I 3.7^a	1,2
pBF121	pBR322/*Sal*I/*Bam*HI	*Sal*I/*Pst*I 0.9 and *Pst*I/*Xho*I 0.5^a	2
pBF122	pBR322/*Sal*I/*Bam*HI	*Sal*I/*Pst*I 0.9 and *Pst*I/*Xho*I 1.5^a	2
pBF124	pBR322/*Pst*I	*Pst*I 1.0	2
pBF125	pBR322/*Pst*I	*Pst*I 1.3	2
pBF126	pBR322/*Pst*I	*Pst*I 1.3	2
pBF130	pKO1/*Sma*I	*Pvu*II 0.7	2
pBF131	pKO1/*Sma*I	*Pvu*II/*Pst*I 0.19 and *Pst*I/*Pvu*II 0.4	2
pBF134	pUC8/*Sal*I/*Bam*HI	*Sal*I/*Xho*I 3.7^a	2
pBF135	pUC8/*Sal*I/*Hind*III	*Sal*I/*Xho*I 3.7^b	2
pBF136	pUC8/*Pst*I	*Pst*I 1.3	2
pBF137	pUC8/*Pst*I	*Pst*I 1.3	2
pBF138	pUC8/*Sma*I	*Sma*I/*Aha*III 1.1	2
pBF139	pUC8/*Sma*I/*Sal*I	*Aha*III/*Sal*I 1.4	2
pBF140	pUC8/*Sma*I	*Aha*III 1.5	2
pBF141	pUC9/*Sma*I	*Aha*III 1.5	2
pBF142	pUC8/*Sma*I/*Bam*HI	*Hpa*I/*Xho*I 1.4^a	2
pBF148	pBR328/*Sal*I/*Eco*RI	*Sal*I/*Aha*III 1.4^c	2
pBF149	pBR328/*Eco*RI/*Bam*HI	*Sma*I/*Aha*III 1.1^d	2

(a) The *Bam*HI site from the vector of pBF111 was included at one end.
(b) The *Hind*III site from the vector of pBF111 was included at one end.
(c) The *Eco*RI site from the vector of pBF139 was included at one end.
(d) The *Eco*RI and *Bam*HI sites from the vector of pBF138 were included at the ends.

TABLE VIII.2

Surface exclusion index of pED208 recombinants

Donor	Recipient	Surface Exclusion Index
pED208/JC6256	pBF101/JM83	15.0
pED208/JC6256	pBF106/JM83	1.0
pED208/JC6256	pBF111/JM83	7.8
pED208/JC6256	pBF120/JM83	50.6
pED208/JC6256	pBF125/JM83	2.2
pED208/JC6256	pBF126/JM83	31.8
pED208/JC6256	pBF134/JM83	170
pED208/JC6256	pBF136/JM83	1.1
pED208/JC6256	pBF137/JM83	93.2
pED208/JC6256	pBF139/JM83	1.6[a]
pED208/JC6256	pBF140/JM83	9.1
pED208/JC6256	pBF141/JM83	13.9[a]
pED208/JC6256	pRS31/JM83	6.0
pED208/JC6256	pED851/JM83	3.8[b]
F*lac*/JC6256	pRS31/JM83	297
F*lac*/JC6256	pED851/JM83	17.5[b]
F*lac*/JC6256	pBF137/JM83	1.7
F*lac*/JC6256	pBF134/JM83	3.8

(a) Growth rates of bacteria containing these chimeras were 2-3 fold less than other JM83 strains. appeared unhealthy.

(b) pED851 can also function as a donor resulting in lowered exclusion indices.

FIGURE VIII.1

HindIII/XhoI map of the pED208 plasmid modified from Finlay et al. (11) showing regions that contain various pED208 plasmid encoded functions, including surface exclusion (sfx).

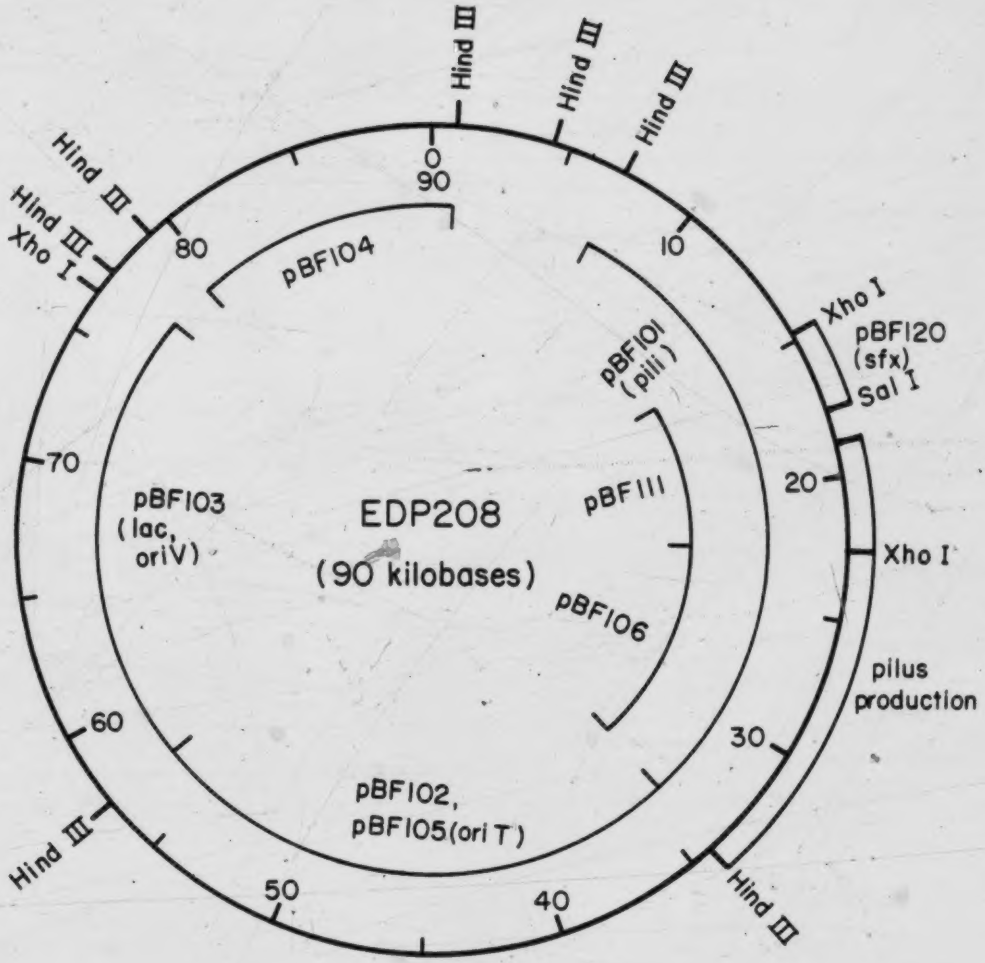

EDP208
(90 kilobases)
Hind III
Hind III
Hind III
Hind III
Hind III
Xho I
Xho I
pBF120
(sfx)
Sal I
Xho I
pilus production
Hind III
Hind III
0
90
10
20
30
40
50
60
70
80
pBF104
pBF101
(pili)
pBF111
pBF106
pBF103
(lac,
oriV)
pBF102,
pBF105(ori T)

FIGURE VIII.2

Restriction enzyme map of the _Sal_I/_Xho_I 3.7 kb fragment of pBF120, with divisions every 500 nucleotides. The location of the two large open reading frames (ORFs), _traS_ and _traT_ are illustrated above the map, with transcription proceeding in the rightward direction. Restriction fragments from pBF120 that were subcloned are shown below the restriction map, with the arrows indicating direction of vector-promoted transcription. The capability (or lack thereof) of excluding pED208 is indicated to the right of the figure.

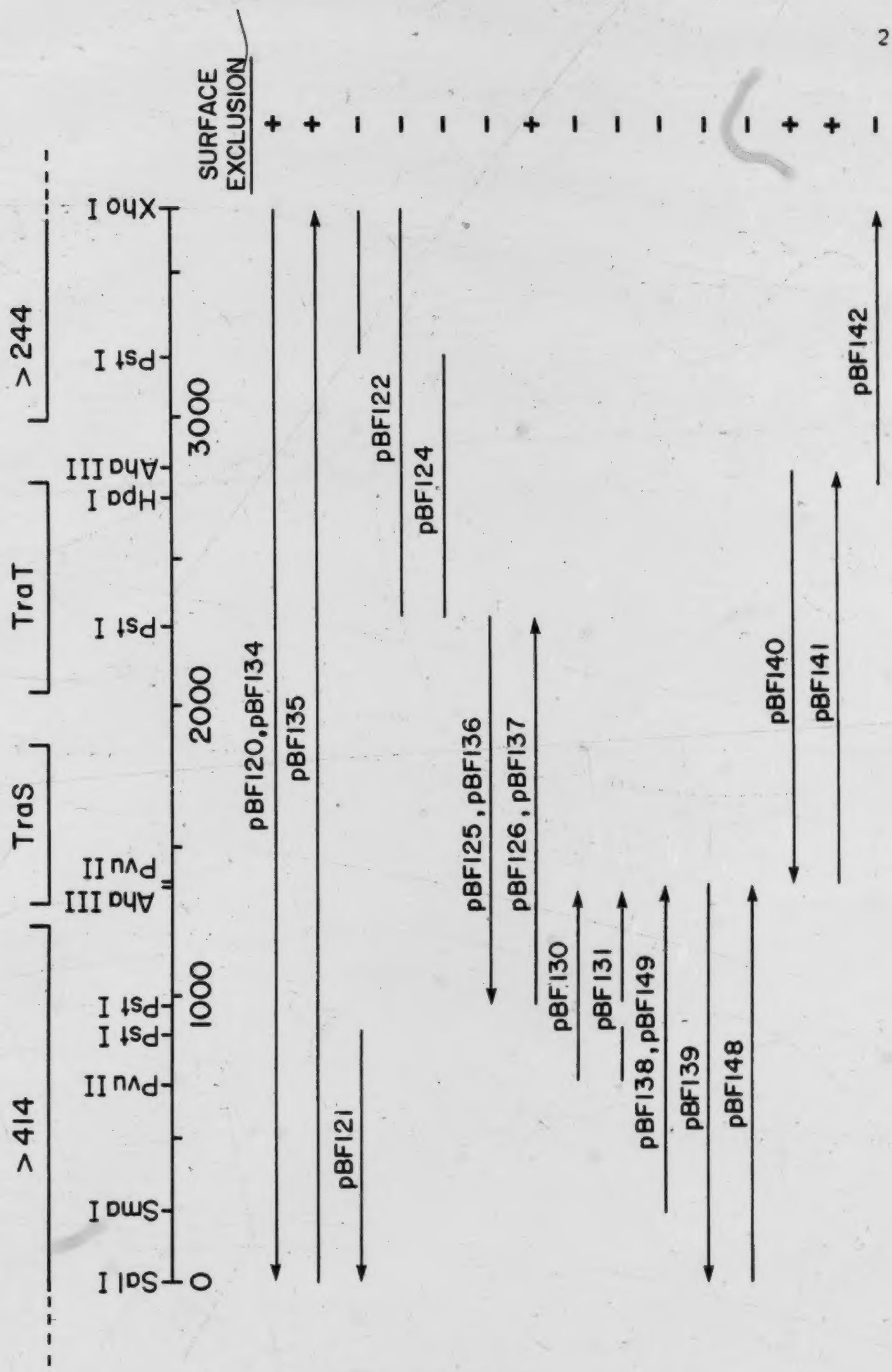

SURFACE EXCLUSION
+
+
-
-
-
-
+
-
-
-
-
-
+
+
-
>244
TraT
TraS
>414
Xho I
Pst I
Aha III
Hpa I
Pst I
Pvu II
Aha III
Pst I
Pst I
Pvu II
Sma I
Sal I
3000
2000
1000
0
pBF120, pBF134
pBF135
pBF121
pBF122
pBF124
pBF125, pBF136
pBF126, pBF137
pBF130
pBF131
pBF138, pBF149
pBF139
pBF148
pBF140
pBF141
pBF142

FIGURE VIII.3

Nucleotide sequence of the 3,725 *Sal*I/*Xho*I fragment, with numbering beginning at the *Sal*I site. *TraS*, *traT* and the two large open reading frames are translated below their respective sequences. Putative ribosome binding sites for *traS* and *traT* are underlined and the putative *traT* promoter is marked.

```
SalI     .         .         .         .        50         .         .         .         .       100
TCGACATTGCGCTGGGTAAAACTGCATCCAGCAGTTGGCAGCAGCAGCAACGTGATGCCCTGAGCGAAGCGCAGAGCCTCAGCAACAGTCTGTCACAGAC
  AspIleAlaLeuGlyLysThrAlaSerSerSerTrpGlnGlnGlnArgAspAlaLeuSerGluAlaGlnSerLeuSerAsnSerLeuSerGlnTh
         .         .         .         .       150         .         .         .         .       200
TTCCAGCCTGGCCACCTCGCAGATGAGTCAGTGGAGCCAGCAGAGGGGAAACAGCGACACCACGGTGACAGGCAGCGATGCCTCTGCCAGCAGCAATCTG
rSerSerLeuAlaThrSerGlnMetSerGlnTrpSerGlnGlnArgGlyAsnSerAspThrThrValThrGlySerAspAlaSerAlaSerSerAsnLeu
         .         .         .         .   SmaI250         .         .         .         .       300
ACCAAAGCGCTGAACACCATTCAGTCTATTGGTTCCCGCTATGCCCGGGATGAGAATACCACTCTGGCGGAAGGGATTCGGGCAGCAGCAACGAAGAGCC
ThrLysAlaLeuAsnThrIleGlnSerIleGlySerArgTyrAlaArgAspGluAsnThrThrLeuAlaGluGlyIleArgAlaAlaAlaThrLysSerG
         .         .         .         .       350         .         .         .         .       400
AGGATATGAGTTTTGGCGCCGGGGCCAGTGTGCAGGGCTCTGTTGACTCAGACCGTCAGATTCTCGGCAAAGTTGCCGGCCTTGTGACCGGAGGGAAAGT
lnAspMetSerPheGlyAlaGlyAlaSerValGlnGlySerValAspSerAspArgGlnIleLeuGlyLysValAlaGlyLeuValThrGlyGlyLysVa
         .         .         .         .       450         .         .         .         .       500
CTCAGGGGAAGCGCATCTGAAAGGTGAGTATACGGGCAGAAGCGGCTCCTCGCACGGCACCAGTTCCGATCTGACGCAGCAGGGCGGGACCAGCAAAGGG
lSerGlyGluAlaHisLeuLysGlyGluTyrThrGlyArgSerGlySerSerHisGlyThrSerSerAspLeuThrGlnGlnGlyGlyThrSerLysGly
         .         .         .         .       550         .         .         .         .       600
TTCAGTGCTCAGGAGCTTAAGGATTTCCGCGATGCAATGGACGTGGTCACCTCCAGCCGCGTGACCGACAGTGGCAGCCATACGGACAACGCATCAGCTT
PheSerAlaGlnGluLeuLysAspPheArgAspAlaMetAspValValThrSerSerArgValThrAspSerGlySerHisThrAspAsnAlaSerAlaS
         .         .         .         .       650         .         .         .         .  PvuII700
CCCTGGCGAACCAGTTGTCGTCCACCTTCTCCAGCCTGAAAAGCCAGGCATCGCAGTATAATGATGCGGTTACCCGCAGCCATGAATACTCACAGCTGGC
erLeuAlaAsnGlnLeuSerSerThrPheSerSerLeuLysSerGlnAlaSerGlnTyrAsnAspAlaValThrArgSerHisGluTyrSerGlnLeuAl
         .         .         .         .       750         .         .         .         .       800
CAGCTATGCAGAGAACAACAGTGCCAGTATTAACCAGAACTACTCGCAGGAGTTTGTCGGGTATGTCACCAGTAAGCGTCCGGGTGAAGCTGACCAGTTG
aSerTyrAlaGluAsnAsnSerAlaSerIleAsnGlnAsnTyrSerGlnGluPheValGlyTyrValThrSerLysArgProGlyGluAlaAspGlnLeu
         .         .         .         .       850         .         .         .      PstI       900
CTTTCTGACGCGGCCAGCCCGGAAGTCCGTGCCGAGCGTGACAGACTGGCGCAGGCATTTGTGGAAGACCGCATGAAGCCGCAACTGCTGCAGGAGTTTG
LeuSerAspAlaAlaSerProGluValArgAlaGluArgAspArgLeuAlaGlnAlaPheValGluAspArgMetLysProGlnLeuLeuGlnGluPheG
         .         .         .         .       950         .    PstI.         .         .      1000
AGCAGAACAGAGGCCGTACCGGTGAGGGGATGGGAAGTGTGGGAGCTTCCTCTGGTGGCGGTAATCTGCAGAGTGAATATGATACGCAACAGACACAAAT
luGlnAsnArgGlyArgThrGlyGluGlyMetGlySerValGlyAlaSerSerGlyGlyGlyAsnLeuGlnSerGluTyrAspThrGlnGlnThrGlnMe
         .         .         .         .      1050         .         .         .         .      1100
GGAAAACCGGGCCAACGCTGGCGGCGTGAGACAGCAGGATGATATCAGCGGTGATGTGACTGACAGACGCCGTCACGCTGAATCTCAGGTCAATGATGAA
tGluAsnArgAlaAsnAlaGlyGlyValArgGlnGlnAspAspIleSerGlyAspValThrAspArgArgArgHisAlaGluSerGlnValAsnAspGlu
         .         .         .         .      1150         .         .         .         .      1200
AAAAATACTGTGATCTCGTCCGGCAATACACCGAAGAAGAATACAATCGTTTGCAGAGTGTGCATGAACAGGGTAATAAAAACTTTAGTGATGCCAAGAG
LysAsnThrValIleSerSerGlyAsnThrProLysLysAsnThrIleValCysArgValCysMetAsnArgValIleLysThrLeuValMetProArgV
         .         .         .         .      1250         .         .         .         .      1300
TGCTGAAGAAAAAAGGCAAAATGATTGGATCGGTGATAATGGTGTAGGGAGAAAGGAATTGAATGATATCAAAGATAACCTGCAAAAAAAGTTTGACCAG
alLeuLysLysLysGlyLysMetIleGlySerValIleMetVal***
         .       TraS        .         .      1350         .    AhaIIIPvuII    .         .      1400
GATTGACAGGAGGGATAATGAAACGTTCGGATCTGGAAAAGGATGCTGCATATATTCTGGATAAGTTTAAACAGCTGGACTATGAAATTCCATCAAACCG
                 MetLysArgSerAspLeuGluLysAspAlaAlaTyrIleLeuAspLysPheLysGlnLeuAspTyrGluIleProSerAsnAr
         .         .         .         .      1450         .         .         .         .      1500
CCAAATTCTAAAGGTTATTTTTTATAAGTTGGTGGTTGTTTATGTATTTCAGTTGCTGTTTATTATCTTTGATATTTTTATAAATTCAAATGTTAGGGAT
gGlnIleLeuLysValIlePheTyrLysLeuValValValTyrValPheGlnLeuLeuPheIleIlePheAspIlePheIleAsnSerAsnValArgAsp
         .         .         .         .      1550         .         .         .         .      1600
TATCATTACTTTGATACATTTGTCATTACTCTGGGATCTAATGCTTTTTTTAGCCTTGTTTTTTTAATGTCAACATACAATCTTGTTAGCCTGAAAATTT
TyrHisTyrPheAspThrPheValIleThrLeuGlySerAsnAlaPhePheSerLeuValPheLeuMetSerThrTyrAsnLeuValSerLeuLysIleS
         .         .         .         .      1650         .         .         .         .      1700
CTTTAGGAAGCGAAATCACAGAACAATCGGTTTTGTTAAAATTGGTTGAGAGAAAAATTAACTCTTATGGTCAATTTCTCATGGTAGTGAATGCTATTGT
erLeuGlySerGluIleThrGluGlnSerValLeuLeuLysLeuValGluArgLysIleAsnSerTyrGlyGlnPheLeuMetValValAsnAlaIleVa
         .         .         .         .      1750         .         .         .         .      1800
TGGATGTGTTTTACTGTCATCGGGGGAGCGGTTTGTAGCTGGACTAGGATTCTCATGGTTTGTAACGTATCTAATTAGCATGATAACGCTACAAACATCT
lGlyCysValLeuLeuSerSerGlyGluArgPheValAlaGlyLeuGlyPheSerTrpPheValThrTyrLeuIleSerMetIleThrLeuGlnThrSer
         .         .         .         .      1850         .         .         .         .      1900
CTTTCTCGATATATGACTCCTGCGGTGGTCAGCAGTTTGTCAAAAGTGAAAGAGTTGCTTTCTGCCTCCCCAAAATAAGCGTCTCATTGTGTATTAAAAC
LeuSerArgTyrMetThrProAlaValValSerSerLeuSerLysValLysGluLeuLeuSerAlaSerProLys***
         .         .         .   -35   .      1950     -10          .         .         .      2000
TTCTGAAAATTCGCACAGTATAACCATAGTGTTGATGTGGTTGTGGTTCTGCCTGTACATTAACTATACGCACATTGAGTGTGTGGACTAATAACAATAG

         .         .         .         .      2050   TraT .         .         .         .      2100
TACCTCCGGGCATGGCATATGCCATGCGACTTATATGGAAATGAGAAAAACTATATGAAACATAACGTTAAGCTGATGGCAATGACTGCCGTTTTATCCT
                                                      MetLysHisAsnValLysLeuMetAlaMetThrAlaValLeuSerS
         .         .         .         .      2150         .         .         .         .      2200
CTGTCCTCGTGCTCTCCGGGTGTGGCGCGATGAGCACAGCGATTAAGAAACGTAACCTCGAAGTCAAAACTCAGATGAGTGAAACCATCTGGCTGGAGCC
erValLeuValLeuSerGLyCysGlyAlaMetSerThrAlaIleLysLysArgAsnLeuGluValLysThrGlnMetSerGluThrIleTrpLeuGluPr
         .         .         .         .      2250         . PstI    .         .         .      2300
GTCCTCACAGAAAACCGTGTATCTCCAGATAAAAAATACCTCTGACAAAGACATGTCTGGCCTGCAGGCGAAAGTAACCAAAG[illegible]GTTCAGGATAAGGGC
oSerSerGlnLysThrValTyrLeuGlnIleLysAsnThrSerAspLysAspMetSerGlyLeuGlnAlaLysValThrLysA[illegible]lGlnAspLysGly
         .         .         .         .      2350         .         .         .         .      2400
TATACCATCACGTCATCGCCGGACAGCGCACACTACTGGATTCAGGCCAATGTCCTGAAAGCCGATAAAATGGACCTGCGTACCG[illegible]CAGGGGTTCCTGA
TyrThrIleThrSerSerProAspSerAlaHisTyrTrpIleGlnAlaAsnValLeuLysAlaAspLysMetAspLeuArgThrAl[illegible]nGlyPheLeuS
         .         .         .         .      2450         .         .         .         .      2500
GCCAGGGCTACGAAGGCGCGATAGCGGGTGCTGCATTAGGTGCGGGGATCACCGGTTATAACTCCAGTTCTGCCGGCGCGACATTAGGTGTTGGCCTGGC
erGlnGlyTyrGluGlyAlaIleAlaGlyAlaAlaLeuGlyAlaGlyIleThrGlyTyrAsnSerSerSerAlaGlyAlaThrLeuGlyValGlyLeuAl
         .         .         .         .      2550         .         .         .         .      2600
GGCGGGCCTGGTGGGAATGGCCGCTGACGCGATGGTGGAAGATATCAACTATACGATGGTGACGGATATTCAGATTTCTGAAAAAACCACCGCAAGCGTG
aAlaGlyLeuValGlyMetAlaAlaAspAlaMetValGluAspIleAsnTyrThrMetValThrAspIleGlnIleSerGluLysThrThrAlaSerVal
         .         .         .         .      2650         .         .         .         .      2700
CAGACTGATAATGTCGCCGCCCTGAAACAGGGTACTTCCGGATATAAAGTGCAGACCAGCACGCAGACCGGTAATCAGCATAAATACCAGACCCGCATCG
GlnThrAspAsnValAlaAlaLeuLysGlnGlyThrSerGlyTyrLysValGlnThrSerThrGlnThrGlyAsnGlnHisLysTyrGlnThrArgIleV
         .     HpaI          .         .      2750         .         .         .         .      2800
TGTCATCTGCCAATAAAGTTAACCTGAAATTCGAAGAAGCAAAACCTGTCCTCGAAGATCAGTTGGCGAAATCTGTCGCCAATATTCTGTAATGCACTCA
alSerSerAlaAsnLysValAsnLeuLysPheGluGluAlaLysProValLeuGluAspGlnLeuAlaLysSerValAlaAsnIleLeu***
         .         . AhaIII  .         .      2850         .         .         .         .      2900
GGGGGGCCGCCTGGTCACCCTGTTTTAAATCTCATTTCCTGAAATACATAACACTATTTAATTACATCAAATAATCTGTTAGCGCATGCAGACAGGCCTG

         .         .         .         .      2950         .         .         .         .      3000
TATCCGGCATGTCTGCTGGTTACCTGCTTCCGGGAAAATATAATGAGCCTGAATCCTCGCGACTTAACGCAGGGCGGTCAGGTCGCTTTTATGCGCCTGA
                                                                                          MetArgLeuL
         .         .         .         .      3050         .         .         .         .      3100
AAATGTTCCTGCAAATAAATAACCTGATTTCTTATTACGTTATTATGGGGACGGTATTATTTGGTATTGCCGTCCTGCTGATGCGAATGAGTATTCAGAA
ysMetPheLeuGlnIleAsnAsnLeuIleSerTyrTyrValIleMetGlyThrValLeuPheGlyIleAlaValLeuLeuMetArgMetSerIleGlnAs
         .         .         .         .      3150         .         .         .         .      3200
TCTGACCAACGGTATTATTTACTGGTTCGTCCGTGTGATGTCGCCGTTTACGGAAAGAATGGTCAGCCAGCCGGTTTATAATATCCGGTATTACGAACAT
nLeuThrAsnGlyIleIleTyrTrpPheValArgValMetSerProPheThrGluArgMetValSerGlnProValTyrAsnIleArgTyrTyrGluHis
  PstI   .         .         .         .      3250         .         .         .         .      3300
ACACTGCAGTACAGCGCCCGGCAGATCCTCAGCGACAATTACACCGTCTATTGTGGGCAGTTACTTAAACAGGAGCTGGTGATTGCAGGGTGTGCTTCGC
ThrLeuGlnTyrSerAlaArgGlnIleLeuSerAspAsnTyrThrValTyrCysGlyGlnLeuLeuLysGlnGluLeuValIleAlaGlyCysAlaSerL
         .         .         .         .      3350         .         .         .         .      3400
TGCTGGTCGCCTTTGTGGCCACATTTGCCGTGTACTGGTATCTCGGGCGTACCGGCCGTAAGCAGAGCGAGGATGAAATCATCGGGGGGCGCGTCCTCAG
euLeuValAlaPheValAlaThrPheAlaValTyrTrpTyrLeuGlyArgThrGlyArgLysGlnSerGluAspGluIleIleGlyGlyArgValLeuSe
         .         .         .         .      3450         .         .         .         .      3500
TGAGTCACCCAAAGACGTGGCGCGCCTGCTGAAAAAACGCGGAGAAGCATCCGATATTCGTATCGATGATCTGCCGCTGAAGCTGGATTCTGAAATTCAG
rGluSerProLysAspValAlaArgLeuLeuLysLysArgGlyGluAlaSerAspIleArgIleAspAspLeuProLeuLysLeuAspSerGluIleGln
         .         .         .         .      3550         .         .         .         .      3600
AACTTCGCCATGCACGGGACAGTATCAACCGGTAAATCCACCCTGATGCGTAAAAACCTGAAGCAACTGCGTGATCGTGGTGACCTGGTCATCATCTACG
AsnPheAlaMetHisGlyThrValSerThrGlyLysSerThrLeuMetArgLysAsnLeuLysGlnLeuArgAspArgGlyAspLeuValIleIleTyrA
         .         .         .         .      3650         .         .         .         .      3700
ACAAGGGCTGTACGTTCGTGGAGGATTTCTACGATGAATCCCGCGATGAAGTGCTGAACGCACTGGATACCCGTTGCCCCAACTGGGATTTATGGGAGGA
spLysGlyCysThrPheValGluAspPheTyrAspGluSerArgAspGluValLeuAsnAlaLeuAspThrArgCysProAsnTrpAspLeuTrpGluGl
         .        XhoI
GTGTCGAACGATATCCGAACTCGAG
uCysArgThrIleSerGluLeuGlu
```

FIGURE VIII.4

Predictive analysis of the secondary structure [Chou and Fasman (7)] and local average hydrophilicity [Hopp and Woods (16)] of the _traS_ gene product. The beta-sheet, alpha-helix, beta-turn and hydrophilicity values are derived from hexapeptide averages along the length of the sequence. Only predictive values greater than one are plotted for beta-sheet, alpha-helix, and beta-turn assignments. Hydrophilic regions appear as positive peaks above zero.

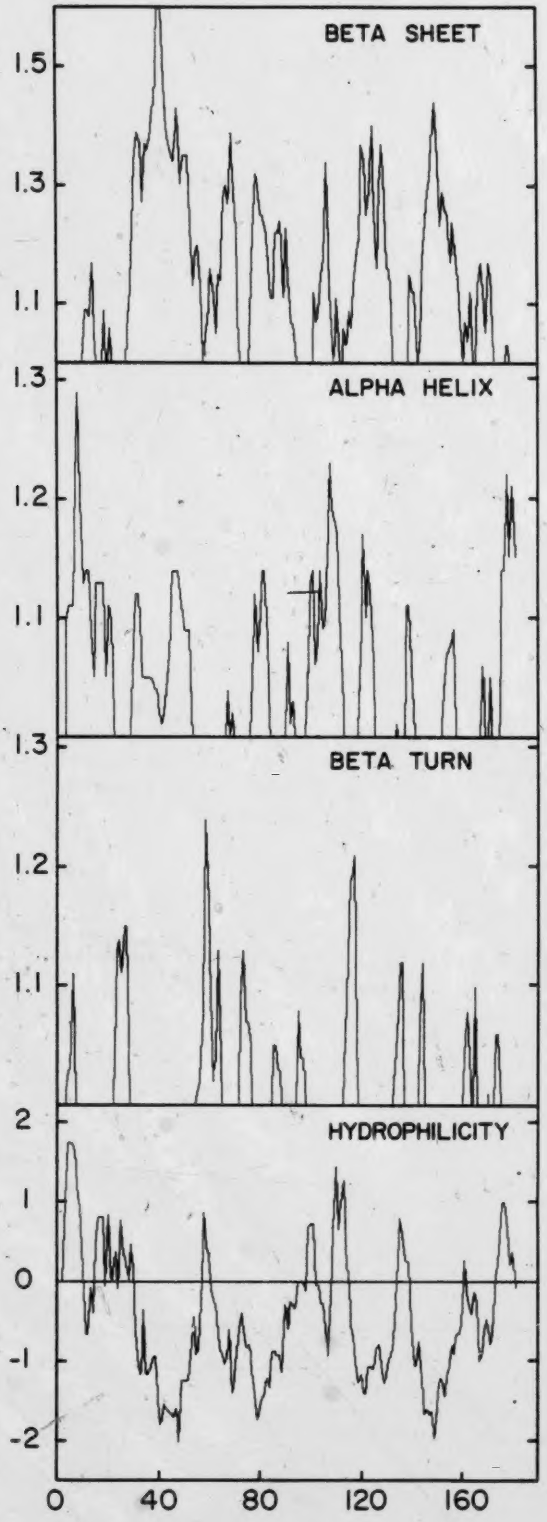

BETA SHEET
1.5
1.3
1.1
ALPHA HELIX
1.3
1.2
1.1
BETA TURN
1.3
1.2
1.1
HYDROPHILICITY
2
1
0
-1
-2
0
40
80
120
160

FIGURE VIII.5

Autoradiograph of a 15% acrylamide gel cotaining [^{35}S]-labelled minicells of pBF140 (lane 1), pBF141 (lane 2), pBF120 (lane 3) and pUC8 (lane 4). The unlabelled arrows represent the mobility of size standards of 43, 25.7, 18.4, 14.3, 12.3, 6.2, and 3 Kdaltons.

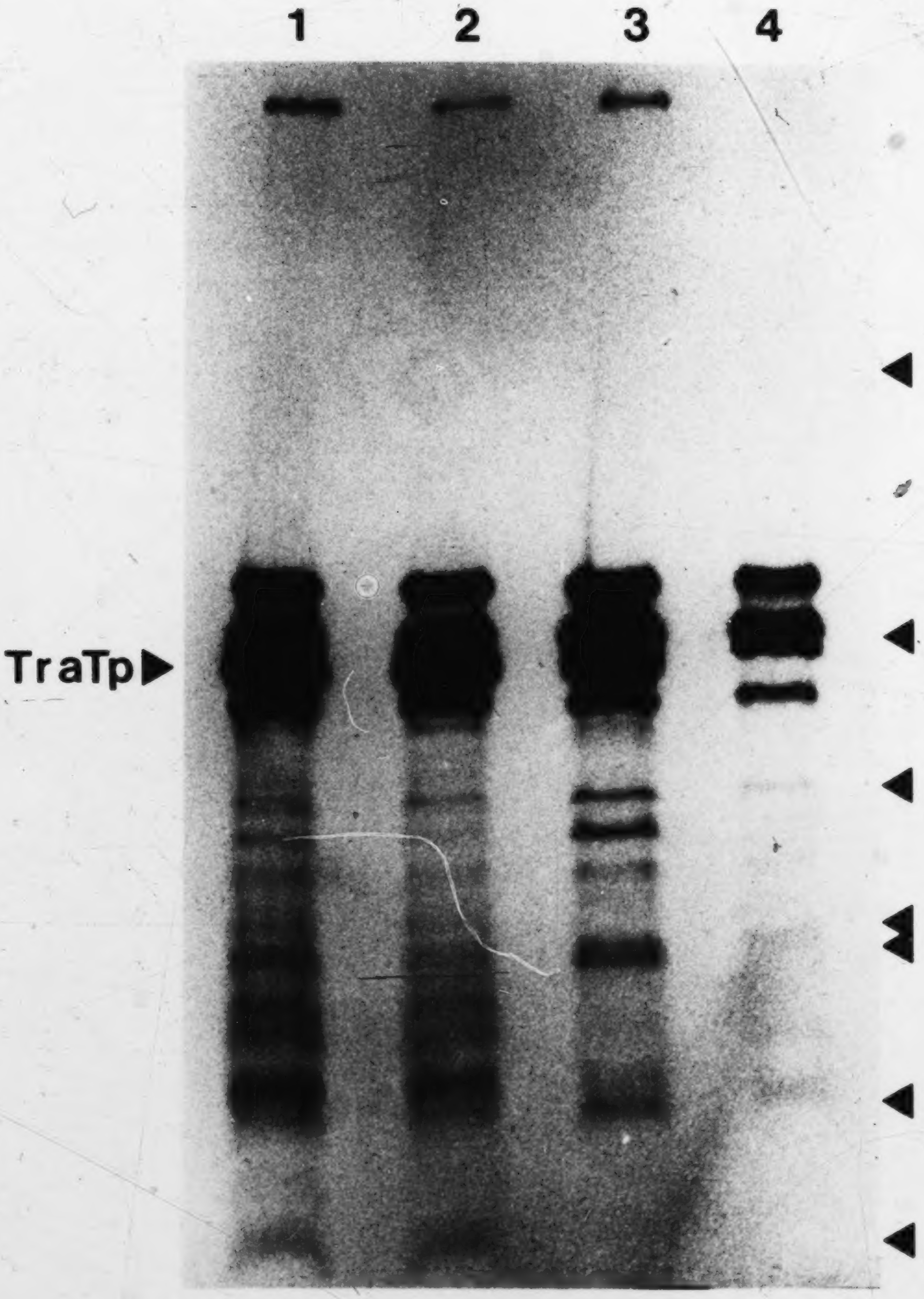

1
2
3
4
TraTp

FIGURE VIII.6

Comparison of TraTps from pED208 (first line) and R-100 (second line). The three bracketed residues at positions 0, 3, and 142 represent the three amino acids found in F TraTp that are different (or missing) in R-100 TraTp (E. Minkley, personal communication). The arrow between residues 22 and 23 represents the proposed signal cleavage site, based on analogy to the F and R-100 systems. Non-homologous residues are boxed, and spaces were inserted to allow maximum alignment. The R-100 sequence is from Ogata _et_ _al_. (24).

```
                                          10                                      20
      Met Lys His Asn Val Lys Leu Met Ala Met Thr Ala Val Leu Ser Ser Val Leu Val Leu
      Met Lys Met Lys     Lys Leu Met     Met Val Ala Leu Val Ser Ser Thr Leu Ala Leu
[Met]         [Thr]
                                          30                                      40
      Ser Gly▼Cys Gly Ala Met Ser Thr Ala Ile Lys Lys Arg Asn Leu Glu Val Lys Thr Gln
      Ser Gly Cys Gly Ala Met Ser Thr Ala Ile Lys Lys Arg Asn Leu Glu Val Lys Thr Gln

                                          50                                      60
      Met Ser Glu Thr Ile Trp Leu Glu Pro Ser Ser Gln Lys Thr Val Tyr Leu Gln Ile Lys
      Met Ser Glu Thr Ile Trp Leu Glu Pro Ala Ser Glu Arg Thr Val Phe Leu Gln Ile Lys

                                          70                                      80
      Asn Thr Ser Asp Lys Asp Met Ser Gly Leu Gln Ala Lys Val Thr Lys Ala Val Gln Asp
      Asn Thr Ser Asp Lys Asp Met Ser Gly Leu Gln Gly Lys Ile Ala Asp Ala Val Lys Ala

                                          90                                      100
      Lys Gly Tyr Thr Ile Thr Ser Ser Pro Asp Ser Ala His Tyr Trp Ile Gln Ala Asn Val
      Lys Gly Tyr Gln Val Val Thr Ser Pro Asp Lys Ala Tyr Tyr Trp Ile Gln Ala Asn Val

                                          110                                     120
      Leu Lys Ala Asp Lys Met Asp Leu Arg Thr Ala Gln Gly Phe Leu Ser Gln Gly Tyr Glu
      Leu Lys Ala Asp Lys Met Asp Leu Arg Glu Ser Gln Gly Trp Leu Asn Arg Gly Tyr Glu

                                          130                                     140
      Gly Ala Ile Ala Gly Ala Ala Leu Gly Ala Gly Ile Thr Gly Tyr Asn Ser Ser Ser Ala
      Gly Ala Ala Val Gly Ala Ala Leu Gly Ala Gly Ile Thr Gly Tyr Asn Ser Asn Ser Ala

                                          150                                     160
      Gly Ala Thr Leu Gly Val Gly Leu Ala Ala Gly Leu Val Gly Met Ala Ala Asp Ala Met
      Gly Ala Thr Leu Gly Val Gly Leu Ala Ala Gly Leu Val Gly Met Ala Ala Asp Ala Met
         [Gly]
                                          170                                     180
      Val Glu Asp Ile Asn Tyr Thr Met Val Thr Asp Ile Gln Ile Ser Glu Lys Thr Thr Ala
      Val Glu Asp Val Asn Tyr Thr Met Ile Thr Asp Val Gln Ile Ala Glu Arg Thr Lys Ala

                                          190                                     200
      Ser Val Gln Thr Asp Asn Val Ala Ala Leu Lys Gln Gly Thr Ser Gly Tyr Lys Val Gln
      Thr Val Thr Thr Asp Asn Val Ala Ala Leu Arg Gln Gly Thr Ser Gly Ala Lys Ile Gln

                                          210                                     220
      Thr Ser Thr Gln Thr Gly Asn Gln His Lys Tyr Gln Thr Arg Ile Val Ser Ser Ala Asn
      Thr Ser Thr Glu Thr Gly Asn Gln His Lys Tyr Gln Thr Arg Val Val Ser Asn Ala Asn

                                          230                                     240
      Lys Val Asn Leu Lys Phe Glu Glu Ala Lys Pro Val Leu Glu Asp Gln Leu Ala Lys Ser
      Lys Val Asn Leu Lys Phe Glu Glu Ala Lys Pro Val Leu Glu Asp Gln Leu Ala Lys Ser

                      245
      Val Ala Asn Ile Leu
      Ile Ala Asn Ile Leu
```

E. BIBLIOGRAPHY

1. **Achtman, M., N. Kennedy, and R. Skurray.** 1977. Cell-cell interactions in conjugating E. coli: role of traT protein in surface exclusion. Proc. Natl. Acad. Sci. U.S.A. **74**:5104-5108.

2. **Achtman, M., P. Manning, C. Edelbluth, and P. Herrlich.** 1979. Export without proteolytic processing of inner and outer membrane proteins encoded by F sex factor tra cistrons in E. coli minicells. Proc. Natl. Acad. Sci. U.S.A. **76**:4837-4841.

3. **Adler, H.I., W.D. Fisher, A. Cohen, and A.A. Hardigree.** 1967. Miniature Escherichia coli cells deficient in DNA. Proc. Natl. Acad. Sci. U.S.A. **57**:321-326.

4. **Aguero, M.E., L. Aron, A.G. Deluca, K.N. Timmins, and F.C. Cabello.** 1984. A plasmid-encoded outer membrane protein, traT, enhances resistance of Escherichia coli to phagocytosis. Infec. Immun. **46**:740-746.

5. **Bradley, D.E.** 1981. Conjugative pili of plasmids in Escherichia coli K-12 and Pseudomonas species, p.217. In S.B. Levy, R.C. Clowes, and E.L. Koening (ed.), Molecular biology, pathogenicity, and ecology of bacterial plasmids. Plenum Publishing Corp., New York.

6. **Bradley, D.E., J.N. Coetzee, T. Bothma, and R.W. Hedges.** 1981. Phage $F_o$lac: an $F_o$lac plasmid-dependent

bacteriophage. J. Gen. Microbiol. 126:405-411.

7. **Chou, P.Y., and Fasman, G.D.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47**:5-148.

8. **Falkow, S., and L.S. Baron.** 1982. Episonic element in a strain of Salmonella typhosa. J. Bacteriol. **84**:581-589.

9. **Ferrazza, D., and S.B. Levy.** 1980. Biosynthesis of a plasmid-encoded outer membrane surface exclusion protein involves processing from a precursor polypeptide. J. Biol. Chem. **255**:8955-8958.

10. **Finlay, B.B., L.S. Frost, and W. Paranchych.** 1984. Localization, cloning and sequence determination of the conjugative plasmid ColB2 pilin gene. J. Bacteriol. **160**:402-407.

11. **Finlay, B.B., W. Paranchych, and S. Falkow.** 1983. Characterization of conjugative plasmid EDP208. J. Bacteriol. **156**:230-235.

12. **Frost, L.S., G.D. Armstrong, B.B. Finlay, B.F.P. Edwards, and W. Paranchych.** 1983. N-terminal amino acid sequencing of EDP208 conjugative pili. J. Bacteriol. **153**:950-954.

13. **Frost, L.S., B.B. Finlay, A. Opgenorth, W. Paranchych and J. Lee.** 1985. Characterization and sequence analysis of pilin from IncF plasmids. J. Bacteriol. **164**:1238-1247.

14. **Frost, L.S., W. Paranchych, and N.S. Willetts.** 1984. DNA sequence of the F traALE region that incudes the gene

for F pilin. J. Bacteriol. 160:395-401.

15. **Hawley, D., and W. McClure**. 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res. 11:2237-2255.

16. **Hopp, T.P., and K.R. Woods**. 1981. Prediction of protein antigenic determinant from amino acid sequences. Proc. Natl. Acad. Sci. USA 78:3824-3828.

17. **Johnson, D.A., and N.S. Willetts**. 1980. Construction and characterization of multicopy plasmids containing the entire F transfer region plasmid. Plasmid 4:292-304.

18. **Kennedy, N., L. Beutin, M. Achtman, R. Skurray, U. Rahmsdorf, and P. Herrlich**. 1977. Conjugation proteins encoded by the F sex factor. Nature 270:580-585.

19. **Lederberg, J., L.L. Cavalli, and E.M. Lederberg**. 1952. Sex compatibility in Escherichia coli. Genetics 37:720-730.

20. **Manning, P.A., and M. Achtman**. 1979. Cell to cell interactions in conjugating Escherichia coli: the involvement of the cell envelope. In M Inouye (ed.), Bacterial outer membrane biogenesis and functions, pp.409-447. New York. Wiley.

21. **Minkley, E.G., and K. Ippen-Ihler**. 1977. Identification of a membrane protein associated with expression of the surface exclusion region of the F tranfer operon. J. Bacteriol. 129:1613-1622.

22. **Minkley, E.G., and N.S. Willetts**. 1984.

Overproduction, purification and characterization of the F traT protein. Mol. Gen. Genet. **196**:225-235.

23. **Moll, A., P.A. Manning, and K.N. Timmis**. 1980. Plasmid-determined resistance to serum bactericidal activity: a major outer membrane protein, the traT gene product, is responsible for plasmid-specified serum resistance in Escherichia coli. Infect. Immun. **28**:359-367.

24. **Ogata, R., C. Winters, and R.P. Levine**. 1982. Nucleotide sequence analysis of the complement resistance gene from plasmid R100. J. Bacteriol. **151**:819-827.

25. **Perumal, N., and E.G. Minkley**. 1984. The product of the F sex factor traT surface exclusion gene is a lipoprotein. J. Biol. Chem. **259**:5357-5360.

26. **Rashtchian, A., J.H. Crooks, and S.B. Levy**. 1983. TraJ independence in expression of traT on F. J. Bacteriol. **154**:1009-1012.

27. **Sanger, F., S. Nicklen, and A.R. Coulson**. 1977. DNA sequencing with chain terminating-inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463-5467.

28. **Scott, J.R.** 1984. Regulation of plasmid replication. Microbiol. Rev. **48**:1-23.

29. **Shine, J., and L. Dalgarno**. 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34-38.

30. **Skurray, R., H. Nagaishi, and A. Clark**. 1976.

Molecular cloning of DNA from F sex factor of E. coli K-12. Proc. Natl. Acad. Sci. U.S.A. **73**:64-68.

31. **Vieira, J., and J. Messing**. 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259-268.

32. **Willetts, N.S.,and J. Maule**. 1980. Characterization of a lambda transducing phage carrying the F conjugation gene traG. Mol. Gen. Genet. **178**:675-680.

33. **Willetts, N.S., and J. Maule**. 1986. Specificities of IncF plasmid conjugation genes. Genet. Res. **47**:1-11.

34. **Willetts, N.S., and R. Skurray**. 1980. The conjugation system of F-like plasmids. Ann. Rev. Genet. **14**:41-76.

35. **Yanisch-Perron, C., J. Vieira, and J. Messing**. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103-119.

Chapter IX

NUCLEOTIDE SEQUENCE OF THE traYALE REGION FROM THE IncFV PLASMID pED208[1]

A. INTRODUCTION

Conjugative plasmids encode the ability to transfer a copy of their genomes to a recipient bacterium. Besides the genes necessary for conjugation, genes encoding toxins, antibiotic resistances, iron metabolism and other pathogenic mechanisms are often found on these transmissible elements. Essential to the conjugative process in F-like plasmids are filamentous appendages on the surface of bacteria called pili, which are encoded within the transfer (*tra*) region of the plasmid (37). Pilin, the structural subunit of pili, is encoded by the *traA* gene of the F plasmid (26). Although there are at least 14 genes involved in F pilus formation, only two other genes, *traQ* and *traG*, have been assigned roles (25). Besides pilus assembly, the *tra* region of F-like conjugative plasmids also encodes genes for conjugal DNA metabolism, surface exclusion and regulation of the

[1] A version of this chapter has been submitted. Finlay, B.B., L.S. Frost, and W. Paranchych. J. Bacteriol.

expression of the transfer operon (37).

pED208 plasmid (formerly called EDP208) is a derepressed derivative of a naturally occurring lac plasmid, Folac, originally identified in Salmonella typhi (8). Cells containing pED208 produce approximately 17 pili per cell, while the naturally derepressed F plasmid produces about 1.4 pili/cell (18). F and pED208 pili are serologically distinct (6) and differ in amino acid composition (3). There appears to be little DNA homology between the transfer regions of pED208 and other F-like plasmids (9) and Folac encodes a system for inhibition of transfer which does not cross-react with the finOP system of these plasmids (36). For these reasons Willetts and Maule (36) have suggested that Folac and pED208 should be included in a separate incompatibility group (IncFolac). However, F and pED208 do share some similarities. Although cells containing pED208 are resistant to F-specific RNA phages, they are sensitive to filamentous DNA phages such as M13 and f1 (5). F and pED208 pilin both contain an acetylated amino terminus (17,19) which plays an important role in the formation of the major antigenic determinants of F and pED208 pili (11,39). Armstrong et al. (3) demonstrated that F and pED208 pilin are similar in size and contain 65 to 70% alpha helix structure and Folkhard et al. (15) have shown that F and pED208 pili are morphologically similar and contain a nearly identical helical array of subunits. Recently we have sequenced the surface exclusion gene, *traT*, from pED208 and found that its product (TraTp) demonstrated a high degree of homology (80%) with the F *traT* product [Chapter VIII, (12)].

In this chapter the nucleotide sequence of *traA*, *traL*, *traE*, and part of *traY* of pED208 plasmid is presented. The amino acid sequences of the predicted products of these genes were compared to the analogous proteins from F plasmid. Also presented is an explanation for the high level of piliation found on cells containing pED208 plasmid.

B. MATERIALS AND METHODS

1. Bacterial strains and plasmids.

Plasmid Fo*lac* (8) was carried in the host strain JE2571 (*Escherichia coli* K-12 *fim fla leu thr lac* Sm^r Nal^r Rif^r) and was kindly donated by D.E. Bradley, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland. pED208 plasmid (formerly EDP208 [13]) was provided by N.S. Willetts, Biotechnology Australia LTD, Roseville, NSW, Australia and was carried in JC6256 (*E. coli* K-12 *trp lac*). All chimeras were transformed into *E. coli* JM83 (34) while *E. coli* P67854 (2) was the strain employed for minicell experiments. Other plasmids employed include pBF101 (13), pBF106 (13) (Fig. 1), and the vectors pUC18 and pUC19 (40). pBF150 contains the 12 kb *Hin*dIII/*Sal*I fragment isolated from pBF106 (map positions 22-34, Fig. 1) inserted into pUC18. pBF151 contains a 2.2 kb *Hin*dIII/*Eco*RI fragment from pED208 in pUC18 (map positions 31.8-34, Fig. 1), while pBF152 consists of a 3.3 kb *Hin*dIII/*Pst*I fragment from pED208 plasmid (positions 30.7-34) inserted into pUC18.

2. Recombinant DNA techniques and DNA sequencing.

Recombinant DNA techniques were performed as described previously (9,18). DNA sequences were obtained by the method of Wallace et al. (35) using reverse and 17 basepair (bp) primers (New England Biolabs, Beverly, Ma.) or 15 bp synthetic oligonucleotides supplied by Regional DNA Synthesis Laboratory, Calgary, Alberta.

3. Isolation and labelling of minicells.

Minicells were purified and labelled with ^{35}S-methionine as described elsewhere (14) and separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) followed by autoradiography (14). To inhibit signal processing (21), 2-phenylethanol was added (final concentration, 0.5%) immediately before the addition of labelled methionine.

C. RESULTS

1. Cloning and Sequencing Strategy.

We have previously reported that a 16.5 kb region of pED208 plasmid DNA encodes the genes necessary for pilus biosynthesis (Fig. IX.1, [13]) and that pED208 DNA does not hybridize with the 1.4 kb *Pst*I fragment of the F plasmid carrying the *traALE* genes (18). By cloning and sequencing the surface exclusion (*sfx*) genes from pED208 (12), we found that these genes (*traS* and *traT*) are separated from the pilus producing region by approximately 2 kb (Fig. IX.1). TraTp, the gene product of pED208 *traT*, shares 80% homology with the TraTps of F and R100 . The position of *traS* and *traT* within the transfer regions of the pED208 and F plasmids are identical. The F *traA* gene is located 20 kb upstream of the *sfx* locus. If pED208 and F contain analogous (or homologous) *tra* operons, it follows that the pED208 *traA* gene should be situated near the *Hin*dIII site at map positon 34 (Fig. IX.1), approximately 18 kb upstream of the *sfx* region.

pBF101 is a chimera containing the 27.5 kb *Hin*dIII fragment from pED208 plasmid (positions 7.5-34) which encodes a functional pilus (13). This indicates that the pilin gene, *traA*, must be located within this *Hin*dIII fragment, to the right of the *Hin*dIII site at position 34 (Fig. IX.1).

A 716 bp *Hin*dIII/*Hpa*I fragment isolated from an IS2 element inserted in R100-1 (Frost, Finlay and Paranchych, in preparation) was used as a probe in Southern blots of digested pED208 DNA to show that this *Hin*dIII site was derived from IS2 (data not shown). The sequence downstream of the *Hin*dIII site in pBF150 was identical to the sequence following the *Hin*dIII site in IS2 (20) in orientation II. Subsequently, two smaller fragments were cloned using the IS2 fragment as a probe. A 2.2 kb *Hin*dIII/*Eco*RI fragment (pBF151) and a 3.3 kb *Hin*dIII/*Pst*I fragment (pBF152) were cloned from pED208 plasmid into pUC18 (see Materials and Methods, Fig. IX.1). Both these fragments contain 870 bp of IS2 DNA upstream of the DNA derived from pED208 (1.4 and 2.5 kb respectively). Sequence analysis of DNA near the *Eco*RI site (map position 31.8 in Fig. IX.1) revealed homology with the F *traE* gene (19). If the genetic ordering of the *tra* operons from pED208 and F are similar, we expected *traY*, *traA* and *traL* to be between the IS2 element and the pED208 *traE* gene. With this in mind we sequenced the entire region between the IS2 element and this *Eco*RI site, and approximately 450 nucleotides downstream of the *Eco*RI site towards the *Pst*I site (map position 30.7). The sequencing strategy employed is presented in Fig. IX.2, and the nucleotide sequence of this region appears in Fig. IX.3.

B. The *traY** gene of pED208.

In the F plasmid, the *traY* gene precedes *traA* (16). Examination of the DNA sequence upstream of pED208 *traA* (whose location is discussed below) revealed a single open reading frame (ORF) encompassing nucleotides -27 to 180. This ORF of 69 residues contained 9 codons derived from the IS2 element and 60 amino acids encoded by pED208 plasmid DNA (Fig. 3). It thus appeared that the IS2 element had inserted into the pED208 *traY* gene, which, nonetheless, produced a functional gene product as evidenced by the fact that pED208 is capable of efficient transfer. Moreover, cells containing the two chimeras pBF101 (Pil$^+$ OriT$^-$) and pBF105 (Pil$^-$ OriT$^+$) are capable of transferring the *oriT*-containing chimera (pBF105) to recipient bacteria (13). By analogy with the F plasmid system, this transfer process presumably requires a functional *traY* protein which is believed to combine with the *traZ* gene product to form an enzyme (*traYZ* endonuclease) which nicks the plasmid DNA at *oriT* (38). The fact that pBF101 is able to mobilize pBF105 indicates that the pED208 *traY* product is functional despite the IS2 insertion.

A potential ribosome binding site (RBS) (32) for the pED208 *traY* gene product (GTGAT) could be assigned within the IS2 element and was followed six nucleotides downstream by an apparent initiation codon for the *traY* protein (position -27, Fig. IX.3). Thus the ORF encompassing positions -27 to 180 could encode a functional traY gene product. The original N-terminus of the pED208 traY protein is presumably upstream of the IS2 element and is currently being sequenced.

Alignment of residues 10 to 60 of the putative pED208 traY* polypeptide with a corresponding region (residues 52-110) of the F traY protein revealed homology levels of 26% (Fig. 4). The same region (residues 10-60) of the pED208 traY* sequence shared 32% homology with the corresponding regions of the R1 traY product (residues 8-75; 10) and 25% homology with the equivalent part of the R100 product (unpublished results). The nine amino acids derived from the end of the IS2 element and fused to the pED208 traY gene would give a traY gene product (69 residues) of approximately the same length as the traY gene products of R1, ColB4 and R100 (75 residues; [10], Chapter VII). Interestingly, the F traY gene encodes the longest gene product (118 residues) of all the F-like traY alleles (10,16). Comparison of all five sequences of traY gene products (pED208, F, R1, R100, ColB4) revealed the most extensive homology to be in the central part of the polypeptide chains, while the greatest variability occurred in the amino- and carboxy-terminal regions.

3. The traA gene of pED208.

Previous biochemical studies in our laboratory have established the sequence of the first 12 amino acid residues and the amino acid composition of pED208 pilin (3,17). An open reading frame began at position 464 which matched this information (Fig. IX.3). The predicted polypeptide contained the same twelve residues that were reported for the N-terminus of pED208 pilin (17), except that Asp-65 Val-66 (Fig. IX.3) are reversed in the reported N-terminal sequence. The amino acid composition predicted by the DNA sequence of pED208 pilin is very similar to the published pilin protein composition if the reported subunit molecular weight is halved from 11,500 to 5750 daltons. The original amino acid composition was based on a molecular weight established by SDS-PAGE. Purified pED208 pilin migrated at a position corresponding to 6,800 daltons if the Laemmli gel system (24) was used (for example, see Fig. IX.5). Neither the predicted protein sequence nor the amino acid analysis of pure pED208 pili contained the amino acids histidine, proline and tryptophan. Finally, the composition of amino acids released by carboxypeptidase A digestion (B. Worobec, unpublished results) corresponded to the last 7 amino acids of the predicted protein sequence. Based on this evidence we concluded that positions 463-654 encoded the processed form of the pED208 pilin protein.

To determine the length of the pilin leader sequence, the ORF upstream of pilin was examined. This ORF originated at position 277 while the first AUG codon occurred at position 298. Nine nucleotides upstream of this codon there is a potential RBS (GGAG; 32). Since there were no other possible translational start sites within this ORF, this was proposed as the start of the traA gene of pED208 plasmid predicting a 55 residue leader sequence in propilin. Immediately before the amino terminus of processed pED208 pilin (Thr-56), there was a sequence that agreed with the consensus sequence for a signal peptidase cleavage site (Ala Ser Ala-55; [30]). Between Arg-37 and Thr-56 there were 18 hydrophobic residues which could constitute a normal signal sequence (4). However, as is the case for the unusually long leader sequence in F propilin (19), the presence of 37 residues preceding this signal sequence remains an enigma.

When the predicted traA polypeptides from F and pED208 were aligned (Fig. IX.4), it was apparent that these two proteins contained regions of significant homology. F plasmid encodes a 51 amino acid leader sequence, four residues shorter than pED208 plasmid. Although there was some homology in these leader sequences (Fig. IX.4), the greatest level of homology occurred within the pilin sequence, with 27 residues being identical (42% homology). Many of the differences found in the pilin sequence arose from hydrophobic amino acid substitutions, although some substitutions resulted in charge differences. Comparison of the pED208 pilin sequence to the four F-like pilin types previously determined (18), revealed similar amounts of homology in all four cases. Moreover, pED208 and R1-19 pilins both terminated with a lysine residue.

4. The traL and traE genes of pED208.

Within the F plasmid transfer operon, traL is positioned immediately downstream of traA (19). Examination of this region in pED208 revealed a large ORF that encoded a polypeptide which is 44% homologous to the F traL gene product (Fig. IX.3 and IX.4), indicating that this was most likely the pED208 traL gene. This polypeptide spanned nucleotides 659 to 961 in Fig. IX.3, and was 99 residues in length (compared with 94 residues for the equivalent F product). Although this gene did not contain a ribosome binding site, it was analogous to the R1 traA-traL system which contains one nucleotide between the stop codon for traA and the start codon for traL (18). Translation of traL in these two systems is probably coupled to that of traA in that both genes utilize the traA ribosome binding site (18).

Twenty nucleotides downstream of the traL stop codon was an ORF of 187 amino acids preceded by a potential RBS 12 nucleotides upstream (CAGG, Fig. IX.3). This predicted protein shared 44% homology with the F traE product (Fig. IX.4) and is probably the pED208 equivalent.

Overlapping the last three codons of traE was an open reading frame that extends beyond the region sequenced (at least 94 amino acids, Fig. IX.3). This ORF was preceded by a strong RBS (GAGGT) five nucleotides upstream (position 1527). The first 24 residues of this predicted polypeptide could function as a leader sequence since there is a lysine residue at position 2, followed by twenty three hydrophobic amino acids. The last four residues of this hydrophobic sequence (Ala Gln Ala Gln-25) could serve as a signal peptidase recognition sequence, with cleavage occurring between Ala-24 and Gln-25. Downstream of residue 24 were stretches of hydrophobic amino acids flanked by charged residues which is characteristic of an integral membrane protein. Genetic evidence indicates that the F traK gene follows traE and minicell experiments have suggested that the traK product may be an outer membrane protein of 24 kd (23). The F plasmid contains an equivalent polypeptide overlapping the last 4 codons in F traE (19). This sequence consists of two positively charged residues followed by 15 hydrophobic amino acids, again remniscent of a signal sequence. Six residues of the F polypeptide can be aligned with the pED208 polypeptide. However, the F system does not contain an obvious RBS as the sequence GGAAACAGA would correspond to the potential RBS, occurring three nucleotides upstream of the proposed initiation codon. Further studies on this region are currently in progress.

5. Polypeptides produced in minicells.

To examine polypeptides produced by the region sequenced, minicells containing the plasmids pUC18, pBF151 or pBF152 were isolated, labelled with ^{35}S-methionine and subjected to SDS-PAGE followed by autoradiography (Fig. IX.5). Minicells containing pBF152 (lane 3) produced polypeptides of 23, 19, 9 and 6.8 kd which were not present in cells carrying pUC18 (lane 1). pBF151 (lanes 2, 5) produced polypeptides of 9 and 6.8 kd in addition to the vector proteins. The presence of the 19 and 23 kd proteins in pBF152 but not pBF151 suggested that these proteins may be either truncated or absent in pBF151. pBF152 contained 1.1 kb of pED208 DNA that was not present in pBF151 and encoded a complete traE gene while pBF151 only encoded the first 125 residues. Based on PAGE results, the reported size of the F traE product is 19,000 daltons (23) while the actual size of the traE products from F and pED208 are 21,200 (19) and 21,300 daltons, respectively. Thus the 19 kd protein produced in minicells containing pBF152 (but not pBF151) may be the traE polypeptide of pED208. The additional DNA contained in pBF152, but not pBF151, was of sufficient length to encode a complete pED208 traK gene, assuming that the pED208 product is similar in size to the 24 kd F traK protein (23). Thus the 23 kd polypeptide may be the pED208 traK product, but confirmation of the assignment of these polypeptides to the traE and traK gene products requires further study.

Minicells containing pBF151 and pBF152 plasmids produced a very intense polypeptide of 6.8 kd while a weaker band migrated with a molecular mass of 9.0 kd (Fig. IX.5). The 6.8 kd protein migrated with the same mobility as purified pED208 pilin, which had a predicted molecular weight of 6850 for an unacetylated product. The addition of 2-phenylethanol (PEA), a known signal peptidase inhibitor (21), caused the intensity of the 6.8 kd protein band to lessen, while the 9 kd band increased in intensity (Fig. IX.5, lanes 4, 5). This was unexpected since pED208 pilin contains a 55 residue leader sequence, corresponding to a propilin molecule of 12,690 daltons in size.

In addition, Laine et al. have demonstrated that at least two additonal tra products, traQ and traG, are required to produce a mature F pilin subunit (25). In the absence of traQ, the propilin polypeptide (14 kd) is produced, which is processed to pilin (7.2 kd) in the presence of traQ and acetylated by the traG gene product. pBF151 encoded neither traQ nor traG based on sequence data and their relative locations in the F tra operon. Immunoblot analysis of cells containing pBF151 revealed that neither the 9 nor 6.8 kd proteins reacted with anti-pilus antibodies (data not shown). However, Worobec et al. (39) have demonstrated that at least 80% of the antibodies raised against pED208 pilin are directed against the acetylated amino-terminus of this molecule. Removing the amino-terminal acetyl group of pED208 pilin results in a 8,300-fold decrease in antibody binding. Laine et al. (25) have demonstrated that the F traG product is necessary for the production of immunologically- active mature pilin protein. If the 6.8 kd polypeptide is an unacetylated, processed pilin molecule, positive detection in an immunoblot is unlikely. Thus it remains a possibility that the 6.8 kd polypeptide is an unacetylated pilin molecule which has been processed without an F-like traQ product. We are unsure of the identity of the 9 kd protein which may be an intermediate product of pilin processing. Alternatively, one of these polypeptides may be the truncated traY* product, which has a predicted molecular weight of 7,590 daltons. There were no other ORFs within pBF151 in either orientation that contained a possible translational start site and were of the correct length to encode these proteins. We are currently isolating these polypeptides and hope to use microsequencing techniques to identify these products.

Although the predicted molecular weight of the traL product is 12,300 daltons, no peptides were visible in the 12 kd range on gels of minicell polypeptides. However, the traL equivalent of F is not visible in minicell preparations either (23).

D. DISCUSSION

Insertion elements are often found in conjugative plasmids (29). IS2 elements have been found within traE of R100-1 (Frost, Finlay and Paranchych, in preparation) and upstream of traT in R100-1 (28). These IS2 elements within the R100 tra operon are both inserted in "orientation II" (20) with respect to the tra operon which is thought to turn on transcription of downstream genes, probably from at least two promoters (22). The IS2 element in pED208 was also found to be in orientation II with respect to the tra genes. Southern analysis of plasmid DNA indicated that there was no IS2 element in the original repressed form of pED208, Folac, while there was one copy in pED208, located in the same EcoRI fragment as the traA gene (data not shown). Since the only apparent difference between these plasmids was the derepressed state of pED208, the IS2 element in pED208 may be the cause of this derepression.

In F-like plasmids finO and finP gene products cause repression of traJ expression (27). The traJ gene product is a regulatory protein which is required for tra operon transcription (33). The absence of either the finO or finP products results in derepression of traJ product synthesis and the concomitant transcription of the traYZ operon. Based on levels of piliation, pED208 is more derepressed than other IncF plasmids. pED208 produces approximately 17 pili/cell, while other derepressed IncF plasmids that are either finO or finP usually produce 1 to 3.6 pili/cell (18). These differences in levels of derepression could be explained by the presence of IS2 near the start of the pED208 tra operon. The normal promoter for the tra operon is the traYZ promoter (the site of traJ activation; [27,33]), located immediately upstream of traY. In pED208, IS2 was located near the beginning of traY. Insertion at this position would uncouple the pED208 tra operon from the traYZ promoter and place it under IS2 transcriptional control, creating the derepressed state seen in cells containing the pED208 plasmid. Since the traY product appeared to be functional, this IS2 element did not block any required pED208 tra functions, but instead provided an alternative tra operon promoter which was not regulated, producing a derepressed state.

Further evidence of IS2-promoter utilization within pED208 comes from transposon-mutagenesis studies (13). We were interested in delineating the region(s) required for pED208 pilus production. By mapping Tn1 and Tn5 insertions, we established a 16.5 kb region required for pilus biosynthesis (13). Many of these Tn::pED208 plasmids contain insertions in the traYALE region. Several map to the right (downstream) of the HindIII site within IS2 (position 34 in Fig. IX.1), and two Tn5 mutations are located in the immediate vicinity of the HindIII site, suggesting that transposon insertions within the IS2 element block pED208 pilus assembly. No Pil⁻ mutations were found that contained insertions further to the left of the HindIII site, the region where finP, traJ and the traYZ promoter map in the F system, indicating that the tra operon of pED208 is appeared to be transcribed independently of traJ/finOP control. pBF101 is a chimera which does not contain the region to the left of HindIII (Fig. IX.1), but includes the remainder of the tra operon and produces pili (13). Although the IS2 promoter(s) that Hinton and Musso (22) mapped to the left of the HindIII site were not present, a promoter in IS2 downstream of the HindIII site or in the vector (pACYC184) presumably provided the required transcriptional activity.

Production of the putative tra products in minicells was unexpected since the vector component of pBF151 and pBF152 (pUC18) did not contain a promoter oriented in the direction required for traALE transcription and the reported promoters in IS2 (22) were not present. This suggests that there was a promoter within the the IS2 element to the right of HindIII which was responsible for transcription of traY and possibly traALE. Since the traY gene product was not visible in minicells (see above), it is possible that the traALE genes were transcribed off a separate promoter which spanned the area of IS2 insertion or was present in the intergenic region between traY and traA where several possible promoter sequences exist (see Fig. IX.3).

The predicted signal sequences of the F and pED208 gene products were both unusual in that they contain 25-30 more residues than usual (4). These two sequences contained only limited homology while the processed pilin proteins had greater similarity (Fig. IX.4). Both leader sequences contained an arginine followed by 18 hydrophobic residues immediately upstream of the signal cleavage sites (residues 37-55 for pED208 plasmid, 33-51 for F plasmid, Fig. IX.4). Frost et al. (19) have proposed that in F propilin the first 11 hydrophobic residues of this region (residues 34-44) could form an alpha-helix which spans the membrane. Following this sequence were seven residues which favored a reverse turn conformation and contained the signal peptidase site. Secondary structure predictions of the pED208 propilin sequence indicated that the equivalent hydrophobic region may be in either alpha-helix or beta-sheet conformation. The reverse turn predicted for F propilin preceding the signal recognition site was not predicted for the pED208 sequence. Nonetheless, based on similarity to the classic leader sequence, this region may span the membrane, orienting the signal recognition site appropriately for signal peptidase cleavage. Although the region upstream of this internal "leader sequence" was not conserved between these two propilin molecules, both contained six positively-charged amino acids. Frost et al. (19) have suggested that these residues may be involved in interactions with the cytoplasmic side of the membrane, allowing the "leader sequence" to span the membrane and positioning the signal peptidase recognition site on the periplasmic side of the inner membrane.

F and pED208 pilin were found to have very similar predicted secondary structures which agreed with the findings that both proteins have 65-70% alpha helix content (3) and have similar X-ray diffraction patterns (15). Most of the length differences between F and pED208 pilin appeared to be due to a shift in the signal cleavage site (Fig. IX.4), resulting in the amino terminus of pED208 pilin being five residues shorter than the equivalent F sequence. The carboxyl terminus of pED208 pilin was one residue shorter than this region in F pilin. It has been shown previously that the amino and carboxyl termini of processed pilin from F-like plasmids vary in length as well as in sequence (18) and that these differences define the specificity of pilus interaction with other cells and phage (36). Thus the differences in length at the ends of pED208 pilin occurred in areas that could tolerate these changes and the differences in sequence were probably responsible for the type-specific characteristics of pED208 pili. The core portions of F and pED208 pilin are both highly hydrophobic in nature and contain many conservative substitutions. This portion is probably important in subunit-subunit interactions within the assembled pilus.

Both F and pED208 confer sensitivity to pilus-specific DNA bacteriophages such as f1 (5). These filamentous phages adhere to the tips of IncF pili, indicating a shared phage-binding domain between these various pilus proteins. Examination of Fig. IX.4 for homologous regions between pED208 and F pilin proteins indicates that the regions of highest homology occur near the amino termini; residues 66-74 of pED208 propilin and residues 67-75 of F propilin. This region is, therefore, a possible candidate for the f1 phage binding domain. Also common to both these proteins is an acetylated amino terminus (Ac-N-Thr for pED208 [17]; Ac-N-Ala for F [19]). However, attempts to block f1 phage attachment with synthetic peptides corresponding to the acetyl group and amino terminal portions of F and pED208 pilin were unsuccessful (unpublished results).

In the F plasmid, traL and traE encode proteins that are required for pilus assembly, but their precise function is unknown. The product of traL is found in the outer membrane while traE product is located in the inner membrane (23). Neither of these membrane proteins appear to have a leader sequence (1). F and pED208 traL and traE products shared about 44% homology, as illustrated in Fig. IX.4. The pED208 products had no recognizable leader sequences, with charged residues distributed throughout the proteins. Secondary structure predictions for traL and traE products from F and pED208 plasmids were similar, indicating that the pED208 proteins may be in the conformations proposed for the F proteins (19). Both F and pED208 traL and traE proteins contained hydrophobic stretches which had high alpha helix and beta sheet potentials and were of sufficient length to span a membrane. Between these regions were charged residues specifying a turn, again characteristic of membrane proteins. Presumably the pED208 and F traL and traE products would be found in similar cellular locations, based on their apparent homology.

When chimeras containing the pED208 tra genes (pBF106, pBF111) were transformed into cells containing point mutations in F traA, traE, traK, traB, traC, traF, traH, traG, or traD, the pED208 gene products were unable to complement the F mutations (data not shown). However, when the chimera pBF101, which encodes the genes necessary for pilus assembly (13), was present with these F tra mutations, all of the F tra plasmids were mobilized into a recipient cell. This was not the case with an F traJ mutation which was not mobilized, suggesting that traJ is required for expression of gene products involved in the physical transfer of the F plasmid (for example, traY product). This data demonstrated that the differences between F and pED208 tra genes were sufficient to prevent any complementation of specific gene products between these two systems. However, the F transfer system was able to utilize a functional pED208 pilus, provided by pBF101, to establish a conjugation bridge between mating cells allowing transfer of the F plasmid to take place. Thus, the F plasmid appeared to be positioned within the cell such that it could take advantage of the mating pair established by the pED208 proteins and the signal for transfer of the F plasmid was triggered.

We have demonstrated that tra genes from pED208 plasmid were more closely related to other IncF tra genes than was originally proposed (36). Although allele differences within the IncFI-IV tra genes were not as great as the differences between these genes and the equivalent pED208 genes, many residues of these gene products were conserved as is their order within the traYZ operon. Bradley et al. have placed pED208 (and Folac) in a unique incompatibility group, IncFolac (6). Based on tra operon similarities with other IncF plasmids, we suggest that pED208 and Folac plasmids should be considered IncF plasmids (IncFV), as originally classified by Datta et al. (7).

By examining conserved and altered sequences in IncF plasmids, some of the region(s) involved in characteristic tra protein functions can be elucidated. These comparisons also provide information about the evolution of the _tra_ genes, suggesting that pED208 plasmid diverged before the IncFI-IV plasmids established their allelic specificities. The pED208 _tra_ genes may thus be a "missing link" in conjugative plasmid evolution, increasing our understanding of the many mechanisms involved in the complex biological phenomenon called conjugation.

FIGURE IX.1.

HindIII/XhoI map of pED208 taken from Finlay et al. (13) showing several plasmid-encoded functions that were cloned including oriT, pilus production and surface exclusion. The locations of the IS2 element and several pED208 tra genes are also shown. The area required for pilus production (positions 18 to 34) was defined by transposon mutagenesis (13), and numbered divisions represent kb.

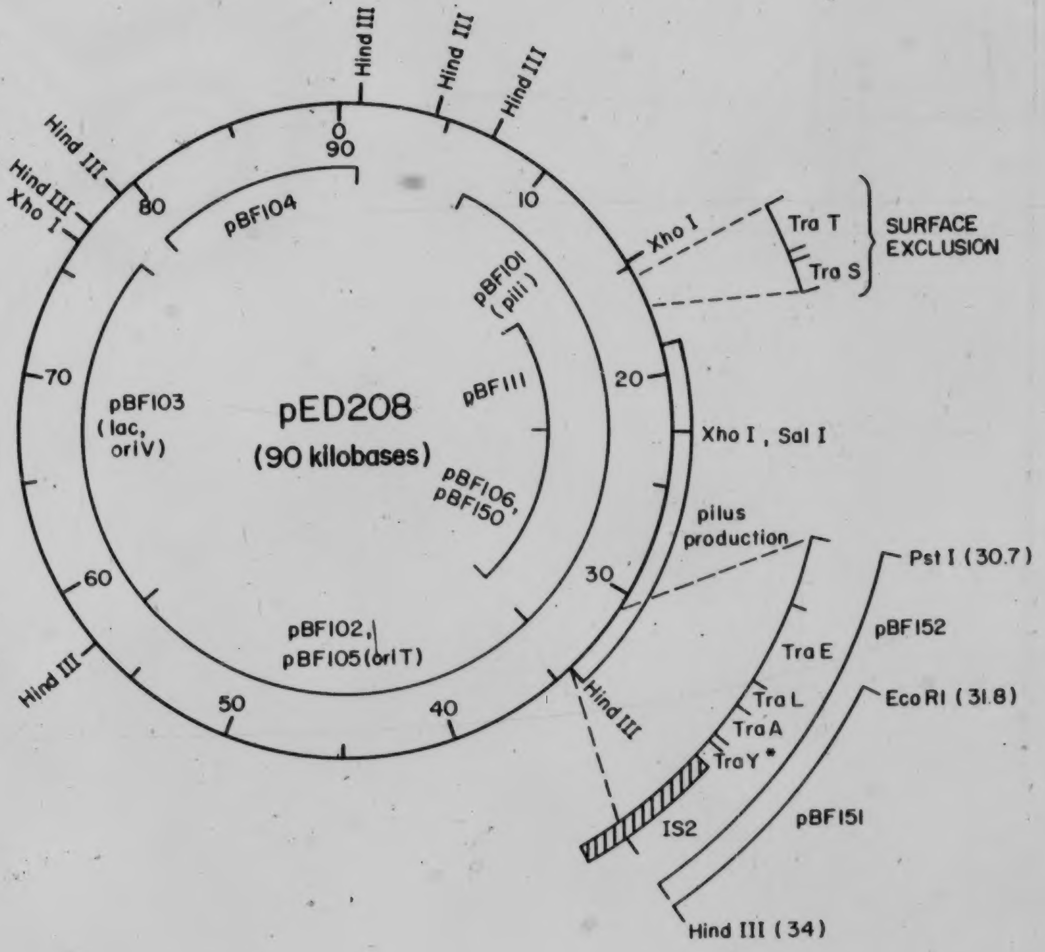

pED208
(90 kilobases)
0
90
10
20
30
40
50
60
70
80
Hind III
Hind III
Hind III
Hind III
Hind III
Xho I
Xho I
Xho I , Sal I
Hind III
Hind III
pBF104
pBF101
(pili)
pBF111
pBF106,
pBF150
pBF103
(lac,
oriV)
pBF102,
pBF105(oriT)
Tra T
Tra S
SURFACE
EXCLUSION
pilus
production
Pst I (30.7)
pBF152
Eco RI (31.8)
Tra E
Tra L
Tra A
Tra Y *
IS2
pBF151
Hind III (34)

FIGURE IX.2.

Restriction map and relative positions of pED208 traYALE. TraY* represents the truncated traY gene that contains an IS2 element. The location of the open reading frame downstream of traE is also shown. The DNA sequencing strategy is included, with tailed arrows representing sequences obtained using custom-made synthtic oligonucleotides and non-tailed arrows represent sequences determined using conventional M13 primers (Materials and Methods). Numbered divisions represent bp and are the same as those used in Fig. IX.3.

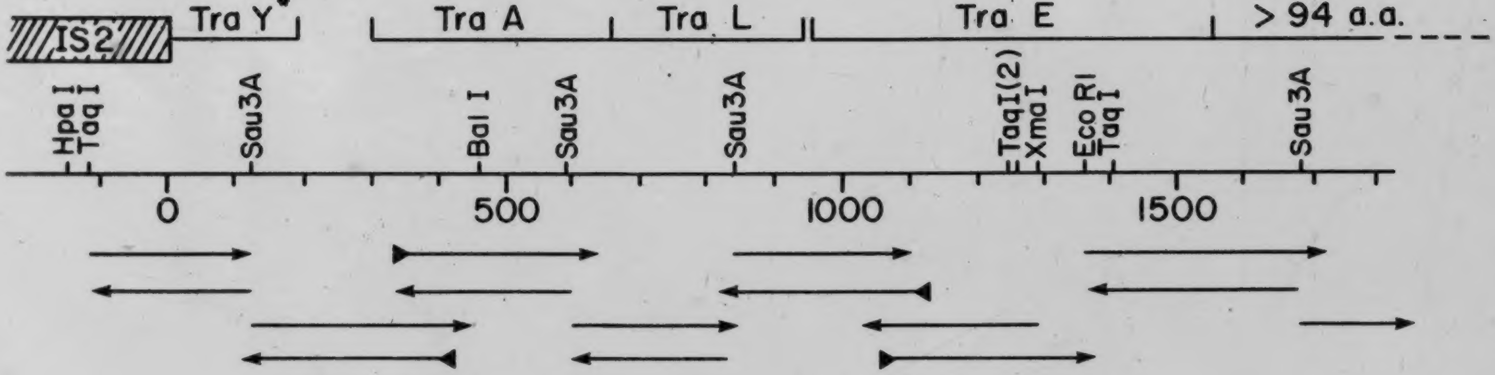

IS2
Tra Y*
Tra A
Tra L
Tra E
> 94 a.a.
Hpa I
Taq I
Sau3A
Bal I
Sau3A
Sau3A
TaqI(2)
Xma I
Eco RI
Taq I
Sau3A
0
500
1000
1500

FIGURE IX.3

Nucleotide and translated sequence of traA, traL, traE and part of traY. Also included are the last 100 nucleotides of IS2 (postions -100 to -1; [20]). Possible ribosome binding sites are underlined, and the vertical arrow represents the processing site of traA product.

```
-100     .         .         .         .        -50        .         .         .         .   ← IS2
TAGTGCGCTGGGTTATCGCTCGCCACGGGAATATCTGCGGCAGCGGGCTTGTAATGGGTTAAGTGATAACAGATGTCTGGAAATATAGGGGCAAATCCA
                                                                  ***     MetSerGlyAsnIleGlyAlaAsnPro

TraY* →  .         .         .         .         50        .         .         .         .        100
ATTGGCAAGAAAGTGAATATTAGCTGTAGTCTGGATGAAGCCATTGATGAACTATTAATGGAGAGTGCACTCAAGAGTGGGTGGAGTAAAAAAAGAGAAG
IleGlyLysLysValAsnIleSerCysSerLeuAspGluAlaIleAspGluLeuLeuMetGluSerAlaLeuLysSerGlyTrpSerLysLysArgGluA
 1                          10                           20                            30

         .     Sau3A       .         .        150        .         .         .         .        200
CTGAATTAAGACTGGAGGATCATTTACGAAGATTTTCTCTCGTACCTGCTGAAGAACAGTATATTGAAAAAAAAGTGGATTAATACGTTAAGTATTCTGT
laGluLeuArgLeuGluAspHisLeuArgArgPheSerLeuValProAlaGluGluGlnTyrIleGluLysLysValAsp***
                 40                            50                            60

         .         .         .         .        250        .         .         .         .       TraA
TTCAGAATCATTGGCCTTATCTGATGGTTCTGTGCCGGATTGACTTAACCCGTCAATACCGCTGGGTGTTAATTGAAATAGCATAAATGGAGTTTTTATG
                                                                                                  Met
                                                                                                    1

         .         .         .         .        350        .         .         .         .        400
AATTTATCCTTTGCAAAAGGCGGCCTCCCTGCGCCTGTAAAAAACCGTGCATGGCAGTACTGCCAGATGGCATGGCGCGGTGTGACCAGTAAAAAAGCGC
AsnLeuSerPheAlaLysGlyGlyLeuProAlaProValLysAsnArgAlaTrpGlnTyrCysGlnMetAlaTrpArgGlyValThrSerLysLysAlaL
                        10                           20                            30

         .         .         .         .        450  BalI      .         .         .         .        500
TGTCCCGTCTGGCTGCGCTGTCTCCGCTGCTGTTACTCGGTGTGGGACAGATGGCCAGTGCAACCGACCTGCTGGCCGGGGGCAAGGATGATGTGAAAGC
euSerArgLeuAlaAlaLeuSerProLeuLeuLeuLeuGlyValGlyGlnMetAlaSerAlaThrAspLeuLeuAlaGlyGlyLysAspAspValLysAl
               40                          50             ↑               60

         .         .         .         .        550        .         .         .         .     Sau3A
CACCTTCGGTGCCGACTCATTCGTCATGATGTGTATCATCATTGCCGAACTGATTGTCGGTGTGGCGATGTATATCCGCACCAAGAACCTGCTGATCCTG
aThrPheGlyAlaAspSerPheValMetMetCysIleIleIleAlaGluLeuIleValGlyValAlaMetTyrIleArgThrLysAsnLeuLeuIleLeu
     70                            80                         90                            100

         .         .         .         .        650      TraL      .         .         .        700
CTGGGCCTGGTTGTGGTTATCGTCTTCACTACCGTCGGTCTTACCTTCATCAAATGATATGGAGGGAAACGATTTAGACAAATACCGTTTCCCGAAAACC
LeuGlyLeuValValValIleValPheThrThrValGlyLeuThrPheIleLys*** MetGluGlyAsnAspLeuAspLysTyrArgPheProLysThr
                         110                              1                         10

         .         .         .         .        750        .         .         .         .        800
CTGTCAGAGCAGAACCGCATTATTGGACTGCCCCTTGATGAAGCTATCCCCGCAGCGGTTGTTTTGCTGTGGGGGTTTTTCACCAAGAAATATTTATTCA
LeuSerGluGlnAsnArgIleIleGlyLeuProLeuAspGluAlaIleProAlaAlaValValLeuLeuTrpGlyPhePheThrLysLysTyrLeuPheS
                  20                           30                           40

         .         .      . Sau3A  .        850        .         .         .         .        900
GTCTGATTATTGCGGCAGTTATCTGGCAACTGATCAGGGCGGCGAAATGTGGAAAAAGCTCCCGCTGGTTATACAACTGGTGTTATTGGTATCTACCCAC
erLeuIleIleAlaAlaValIleTrpGlnLeuIleArgAlaAlaLysCysGlyLysSerSerArgTrpLeuTyrAsnTrpCysTyrTrpTyrLeuProTh
      50                           60                            70                          80

         .         .         .         .        950        .         .         .  TraE  .       1000
TGAATTATTCCGGGTTGTCTACCGCGTTATTCCCGATTCCAGTTTCAGGAAATGGATAAAATAAATTACCGCAGGGCCATATTATGGAAATTAGCGCTCG
rGluLeuPheArgValValTyrArgValIleProAspSerSerPheArgLysTrpIleLys***                     MetGluIleSerAlaAr
                 90                                    100                               1

         .         .         .         .       1050        .         .         .         .       1100
TAATTCATCCACCAAAATTATTGCCGTTATTATTCTGTCGCTCACGACACTGGTCTTTATTCTGGGCTTTGCCGTGGCATGGCTGGCTGTGACCAACAGA
gAsnSerSerThrLysIleIleAlaValIleIleLeuSerLeuThrThrLeuValPheIleLeuGlyPheAlaValAlaTrpLeuAlaValThrAsnArg
          10                           20                           30

         .         .         .         .       1150        .         .         .         .       1200
GCACTCATCCACGAGCAACGTACAGTCGTCACACCGATGACCTACAACGCGCCTTTTGTGGTATCCGAAACTAAAGCCGATACCGAGTATTTCCGGATGA
AlaLeuIleHisGluGlnArgThrValValThrProMetThrTyrAsnAlaProPheValValSerGluThrLysAlaAspThrGluTyrPheArgMetM
 40                           50                           60                            70

         .         .         .         .        TaqI       .         .       . TaqI    .      XmaI
TGACGCTGTCCTTCCTGGCCCTGCGTCTCAATGTCTCGCCTGAAACGGTCGATTCCAACCACGCTTTTCTGATGTCTTTTGTCGAACCAGAGGCCCGGGA
etThrLeuSerPheLeuAlaLeuArgLeuAsnValSerProGluThrValAspSerAsnHisAlaPheLeuMetSerPheValGluProGluAlaArgGl
                      80                           90                            100

         .         .         .         .       1350    EcoRI     .         .         .       1400
GGAATTTAAAAAAGTCCTCCAGGAAGAGGCTGCACAGATTAAAGCCAATGATGTGAATTCCACGTTCTACACCACAGAAATCAATGTCTATCCCGTTGAC
uGluPheLysLysValLeuGlnGluGluAlaAlaGlnIleLysAlaAsnAspValAsnSerThrPheTyrThrThrGluIleAsnValTyrProValAsp
          110                          120                           130

       TaqI        .         .         .       1450        .         .         .         .       1500
GCCCGCATCGATGTGCGTGGCGTGCTGAAAATGTGGATTGGTAACTCCAGGCCTTCCACGGAAATAAAAACCTACCGGCTTCGTCTGAAGTACACCGGAG
GlyArgIleAspValArgGlyValLeuLysMetTrpIleGlyAsnSerArgProSerThrGluIleLysThrTyrArgLeuArgLeuLysTyrThrGlyG
140                         150                           160                           170

         .         .         .       TraK?     1550        .         .         .         .       1600
GCTTTACCCGTATCGGTCGGTTCTATGAGGTGACGAATGAAAAATAACCTTCCGGCGTTTTTGTTTGGTACAGCCATGATGGTTGTTATGCCCCCGGCTG
                                                   MetLysAsnAsnLeuProAlaPheLeuPheGlyThrAlaMetMetValValMetProProAlaA
lyPheThrArgIleGlyArgPheTyrGluValThrAsnGluLys***                                10                             20
                     180

         .         .         .         .       1650        .         .  Sau3A  .         .       1700
CGCAGGCACAGTCACCTGCCACGATAAGCCTGCCACAGGGCGGCCAGTTCCGGCTGAGTATCAGCAATACCGATCCCAACATGATTTTTATCCCAGGAGA
laGlnAlaGlnSerProAlaThrIleSerLeuProGlnGlyGlyGlnPheArgLeuSerIleSerAsnThrAspProAsnMetIlePheIleProGlyAs
                        30                           40                           50

         .         .         .         .       1750        .         .         .         .       1800
CAAAGTCACCGCCATCACAGCGCCCGGGGGAATGCTGGCAGACAAGCGACTGACAAGGGCTGGAGGTGTGCTGTTCACCTCGGTGGCCACCCGTACTTTC
pLysValThrAlaIleThrAlaProGlyGlyMetLeuAlaAspLysArgLeuThrArgAlaGlyGlyValLeuPheThrSerValAlaThrArgThrPhe
               60                           70                            80

     1810
ACGATTTTTGTAGGAA
ThrIlePheValGlu
   90
```

FIGURE IX.4.

Comparison of pED208 (first line) and F (second line) traYALE open reading frames. Sequences were aligned by the program of Queen and Korn (31), and homologous residues boxed. The vertical arrows represent the processing sites for F and pED208 propilin. The F traY and traALE sequences are from Fowler et al. (16) and Frost et al. (19).

TraY*

```
1   IleGlyLysLysValAsn   IleSerCysSerLeuAspGluAlaIleAspGluLeuLeu
52  LysSerLeu   ValLysTyrLeuThrMetArgLeuAspAspGluThrAsnGlnLeuLeu

20  MetGluSerAlaLeuLysSerGlyTrpSerLysLysArgGluAlaGluLeuArgLeuGlu
71  IleAlaAlaLysAsnArgSerGlyTrpCysLysThrAspGluAlaAlaAspArgValIle

40  AspHisLeuArgArgPheSerLeuValProAlaGluGluGlnTyr      IleGluLys
91  AspHisLeuIleLysPheProAspPheTyrAsnSerGluIlePheArgGluAlaAspLys

58  Lys      Val      Asp
111 GluGluAspIleThrPheAsnThrLeu
```

TraA

```
1   MetAsnLeuSerPheAlaLysGlyGlyLeuProAlaProValLysAsnArgAlaTrpGln
1   MetAsnAlaValLeuSerValGlnGlyAlaSerAlaProValLysLysLysSerPhe

21  TyrCysGlnMetAlaTrpArgGlyValThrSerLysLysAlaLeuSerArg   LeuAla
20  PheSerLysPheThr   Arg   LeuAsnMetLeuArg   LeuAlaArgAlaValIle
                                                               ↓
40  AlaLeuSerProLeuLeuLeu            LeuGlyValGlyGlnMetAlaSerAla
37  ProAlaAlaValLeuMetMetPhePheProGlnLeuAlaMetAlaAlaGlySerSerGly
                                                 ↑
56  ThrAspLeuLeuAlaGlyGlyLysAspAspValLysAlaThrPheGlyAlaAspSerPhe
57  GlnAspLeuMetAlaSerGlyAsnThrThrValLysAlaThrPheGlyLysAspSerSer

76  ValMetMetCysIleIleIleAlaGluLeuIleValGlyValAlaMetTyrIleArgThr
77  ValValLysTrpValValLeuAlaGluValLeuValGlyAlaValMetTyrMetMetThr

96  LysAsnLeuLeuIleLeuLeuGlyLeuValValValIleValPheThrThrValGlyLeu
97  LysAsnValLysPheLeuAlaGlyPheAlaIleIleSerValPheIleAlaValValMet

116 ThrPheIleLys
117 AlaValValGlyLeu
```

TraL

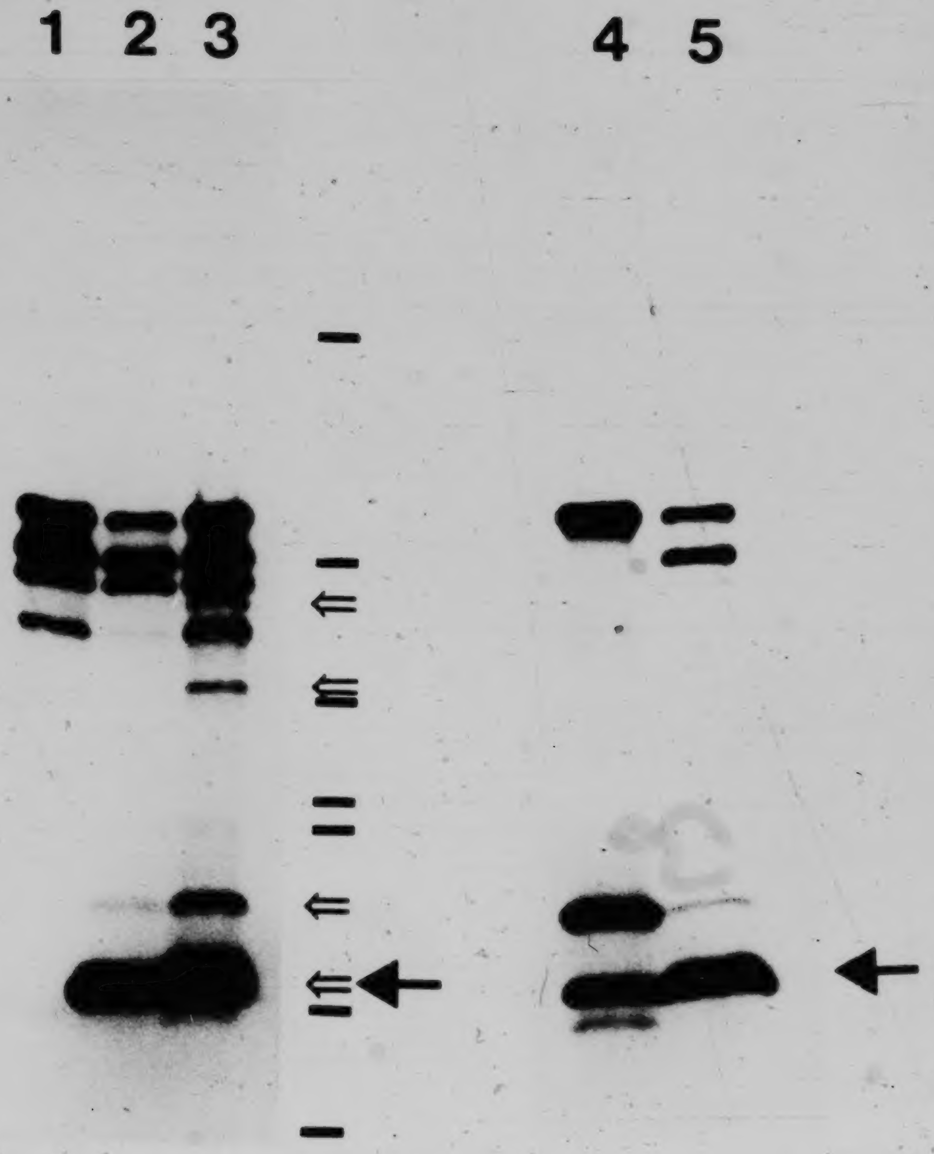
```
1   MetGluGlyAsnAsp      LeuAspLysTyrArgPheProLysThrLeuSerGluGln
1   MetSerGlyAspGluAsnLysLeuLysLysTyrArgPheProGluThrLeuThrAsnGln

19  AsnArgIleIleGlyLeuProLeuAspGluAlaIleProAlaAlaValValLeuLeuTrp
21  SerArgTrpPheGlyLeuProLeuAspGluLeuIleProAlaAlaIleCysIleGlyTrp

39  GlyPhePheThrLysLysTyrLeuPheSerLeuIleIleAlaAlaValIleTrpGlnLeu
41  GlyIleThrThrSerLysTyrLeuPheGlyIleGlyAlaAlaValLeuValTyrPheGly

59  IleArgAlaAlaLysCysGlyLysSerSerArgTrpLeuTyrAsnTrpCysTyrTrpTyr
61  IleLysLysLeuLysLysGlyArgGlySerSerTrpLeuArgAspLeuIleTyrTrpTyr

79  LeuProThrGluLeuPheArgValValTyrArgValIleProAspSerSerPheArgLys
81  MetProThrAlaLeu      Leu      ArgGlyIle   Phe   SerTrpIleLys
```

TraE

```
1   MetGluIleSerAlaArgAsnSerSerThrLysIleIleAlaValIleIleLeuSerLeu
1   MetGluHisGlyAlaArgLeuSerThrSerArgValMetAlaIleAlaPheIlePheMet

21  ThrThrLeuValPheIleLeuGlyPheAlaValAlaTrpLeu   AlaValThrAsn
21  SerValLeuIle   ValLeuSerLeuSerValAsnValIleGlnGlyValAsnAsnTyr

39  ArgAlaLeuIleHisGluGlnArgThrValValThrProMetThrTyrAsnAlaProPhe
40  Arg   LeuGlnAsnGluGlnArgThrAlaValThrProMetAlaPheAsnAlaProPhe

59  ValValSerGluThrLysAlaAspThrGluTyrPheArgMetMetThrLeuSerPheLeu
59  AlaValSerGlnAsnSerAlaAsnAlaSerTyrLeuGlnGlnMetAlaLeuSerPheIle

79  AlaLeuArgLeuAsnValSerProGluThrValAspSerAsnHisAlaPheLeuMetSer
79  AlaLeuArgLeuAsnValSerSerGluThrValAspAlaSerHisGlnAlaLeuLeuGln

99  PheValGluProGluAlaArgGluGluPheLysLysValLeuGlnGluGluAlaAlaGln
99  TyrIleArgProGlyAlaGlnAsnGlnMetLysValIleLeuAlaGluGluAlaLysArg

119 IleLysAlaAsnAspValAsnSerThrPheTyrThrThrGluIleAsnValTyrProVal
119 IleLysAsnAspAsnValAsnSerAlaPhePheGlnThrSerValArgVal   CysLeu

139 AspGlyArgIle   AspValArgGlyValLeuLysMetTrpIleGlyAsnSerArgPro
138 SerMetAlaValTrpLysPheArgGlyValLeuLysThrTrpIleGlyAspSerLysPro

158 SerThrGluIleLysThrTyrArgLeuArgLeuLysTyrThrGlyGlyPheThrArgIle
158 PheThrAspIleLysHisTyrIleLeuIleLeuLysArgGluAsnGlyValThrTrpLeu

178 GlyArgPheTyrGluValThrAsnGluLys
178 AspAsnPheGlyGluThrAspAspGluLysLys
```

FIGURE IX.5

Autoradiograph of a 15% polyacrylamide gel containing [^{35}S]-labelled minicells of: 1)pUC18; 2,5)pBF151; 3)pBF152; 4)pBF151 treated with 0.5% phenethyl alcohol. Solid bars represent the mobility of size standards of (from the top) 43, 25.7, 18.4, 14.3, 12.3, 6.2 and 3 kilodaltons. Polypeptides produced by minicells containing pBF152 not present in pUC18-containing minicells are marked by open arrows corresponding to molecular weights of 23, 19, 9 and 6.8 kd. The solid arrows represent the mobility of pED208 pilin (6.8 kd).


1
2
3
4
5

E. BIBLIOGRAPHY

1. **Achtman, M., P. Manning, C. Edelbluth, and P. Herrlich**. 1979. Export without proteolytic processing of inner and outer membrane proteins encoded by F sex factor _tra_ cistrons in _Escherichia coli_ minicells. Proc. Natl. Acad. Sci. U.S.A. **76**:4837-4841.

2. **Adler, H.I., W.D. Fisher, A. Cohen, and A.A. Hardigree**. 1967. Miniature _Escherichia coli_ cells deficient in DNA. Proc. Natl. Acad. Sci. U.S.A. **57**:321-326.

3. **Armstrong, G.D., L.S. Frost, P.A. Sastry, and W. Paranchych**. 1980. Comparative biochemical studies on F and EDP208 conjugative pili. J. Bact. **141**:333-341.

4. **Benson, S.A., M.N. Hall, and T.J. Silhavy**. 1985. Genetic analysis of protein export in _Escherichia coli_ K-12. Annu. Rev. Bioch. **54**:101-134.

5. **Bradley, D.E.** 1981. Conjugative pili of plasmids in _Escherichia coli_ K-12 and _Pseudomonas_ species, p.217. _In_ S.B. Levy, R.C. Clowes, and E.L. Koening (ed.), Molecular biology, pathogenicity, and ecology of bacterial plasmids. Plenum Publishing Corp., New York.

6. **Bradley, D. E. and E. Meynell**. 1978. Serological characteristics of pili determined by the plasmids R711b and

F_lac_. J. Gen. Microbiol. **108**:141-149.

7. **Datta, N.** 1975. Epidemiology and classification of plasmids. In Microbiology 1974, ed. D. Schlessinger, pp 9-15. Am. Soc. Micro., Washington, D.C.

8. **Falkow, S., and L.S. Baron.** 1982. Episomic element in a strain of Salmonella typhosa. J. Bacteriol. **84**:581-589.

9. **Finlay, B.B., L.S. Frost, and W. Paranchych.** 1984. Localization, cloning and sequence determination of the conjugative plasmid ColB2 pilin gene. J. Bacteriol. **160**:402-407.

10. **Finlay, B.B., L.S. Frost, and W. Paranchych.** 1986. Nucleotide sequence of the R1-19 plasmid transfer genes traM, finP, traJ, traY and traYZ promoter. J. Bacteriol. May issue.

11. **Finlay, B.B., L.S. Frost, and W. Paranchych.** 1985. The major antigenic determinant of F and ColB2 pili. J. Bacteriol. **163**:331-335.

12. **Finlay, B.B., and W. Paranchych.** 1986. Nucleotide sequence of the surface exclusion genes traS and traT from the IncFolac plasmid pED208. J. Bacteriol. June issue.

13. **Finlay, B.B., W. Paranchych, and S. Falkow.** 1983. Characterization of conjugative plasmid EDP208. J. Bacteriol. **156**:230-235.

14. **Finlay, B., B. Pasloske, and W. Paranchych.** 1986. Expression of Pseudomonas aeruginosa PAK pilin gene in Escherichia coli. J. Bacteriol. **165**:625-630.

15. **Folkhard, W., K.R. Leonard, S. Malmsey, D.A. Marvin,**

J. Dubochet, A. Engel, M. Achtman and R. Helmuth. 1979. X-ray diffraction and electron microscope studies on the structure of bacterial F pili. J. Mol. Biol. **130**:145-160.

16. **Fowler, T., L. Taylor, and R. Thompson**. 1983. The control region of the F plasmid transfer operon: DNA sequence of the *traJ* and *traY* genes and characterization of the *traYZ* promoter. Gene **26**:79-89.

17. **Frost, L.S., G.D. Armstrong, B.B. Finlay, B.F.P. Edwards, and W. Paranchych**. 1983. N-terminal amino acid sequencing of EDP208 conjugative pili. J. Bacteriol. **153**:950-954.

18. **Frost, L.S., B.B. Finlay, A. Opgenorth, W. Paranchych and J. Lee**. 1985. Characterization and sequence analysis of pilin from IncF plasmids. J. Bacteriol. **164**:1238-1247.

19. **Frost, L.S., W. Paranchych, and N.S. Willetts**. 1984. DNA sequence of the F *traALE* region that incudes the gene for F pilin. J. Bacteriol. **160**:395-401.

20. **Ghosal, D., H. Sommer, and H. Saedler**. 1979. Nucleotide sequence of the transposable DNA-element IS2. Nuc. Acid. Res. **6**:1111-1122.

21. **Halegoua, S., and M.Inouye**. 1979. Translocation and assembly of outer membrane proteins of *Escherichia coli*: Selective accumulation of precursors and novel assembly intermediates caused by phenethyl alcohol. J. Mol. Biol. **130**:39-61.

22. **Hinton, D.M., and R.E. Musso**. 1983. Specific *in vitro*

transcription of the insertion sequence IS2. J. Mol. Biol. **169**:53-81.

23. **Kennedy, N., L. Beutin, M. Achtman, R. Skurray, U. Rahmsdorf, and P. Herrlich**. 1977. Conjugation proteins encoded by the F sex factor. Nature **270**:580-585.

24. **Laemmli, U.K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680-685.

25. **Lainé, S., D. Moore, P. Kathir, and K. Ippen-Ihler**. 1985. Genes and gene products involved in the synthesis of F-pili, p. 535-553. In D. Helinski, S. Cohen, D. Clewell, D. Jackson, and A. Hollaender (eds.), Plasmids in Bacteria. Plenum Publishing Company, New York.

26. **Minkley, E.G., S. Polen, C.C. Brinton and K. Ippen-Ihler**. 1976. Identification of the structural gene for F pilin. J. Mol. Biol. **108**:111-121.

27. **Mullineaux, P., and N. Willetts**. 1984. Promoters in the transfer region of plasmid F. In D. Helinski, S. Cohen, D. Clewell, and A. Hollaender (eds.), Plasmids in Bacteria, pp.605-614, Plenum Press, New York.

28. **Ogata, R., C. Winters, and R.P. Levine**. 1982. Nucleotide sequence analysis of the complement resistance gene from plasmid R100. J. Bacteriol. **151**:819-827.

29. **Ohtsubo, H. and E. Ohtsubo**. 1977. Repeated DNA sequences in plasmids, phages, and bacterial chromosomes. p.49-63. In A.I. Bukhari, J.A. Shapiro and S.L. Adhya

(eds.), DNA Insertion Elements, Plasmids, and Episomes. Cold Spring Harbor Laboratory.

30. **Perlman, D., and H.O. Halverson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391-409.

31. **Queen, C., and L.J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids. Res. **12**:581-599.

32. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34-38.

33. **Thompson, R., and L. Taylor.** 1982. Promoter mapping and DNA sequencing of the F plasmid transfer genes traM and traJ. Mol. Gen. Genet. **188**:513-518.

34. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259-268.

35. **Wallace, R.B., M.J. Johnson, S.V. Suggs, K. Miyoshi, R. Bhatt, and K. Itakura.** 1981. A set of synthetic oligodeoxyribonucleotide primers for DNA sequencing in the plasmids vector pBR322. Gene **16**:21-26.

36. **Willetts, N.S., and J. Maule.** 1986. Specificities of IncF plasmid conjugation genes. Genet. Res. **47**:1-11.

37. **Willetts, N.S., and R. Skurray.** 1980. The conjugation

system of F-like plasmids. Ann. Rev. Genet. 14:41-76.

38. **Willetts, N.S. and B. Wilkins.** 1984. Processing of plasmid DNA during conjugation. Microbiol. Rev. 48:24-41.

39. **Worobec, E. A., W. Paranchych, J. M. R. Parker, A. K. Taneja, and R. S. Hodges.** 1985. Antigen-antibody interaction: the immunodominant region of EDP208 pili. J. Biol. Chem. 260:938-943.

40. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.

Chapter X

DISCUSSION

An overview of the processes involved in conjugation of IncF plasmids was presented in Chapter I. The aim of this chapter is to relate the findings presented in Chapters II-IX to our general understanding of conjugation. This chapter is divided into sections similar to the divisions in Chapter I, allowing comparisons to be made between structure and function of these conjugal alleles.

A. PILUS BIOSYNTHESIS

The traA (pilin) genes of F-like (IncFI-IV) plasmids are strongly conserved (Chapter III, 4). The less related pED208 traA gene encodes several features that are common to F-like pilin genes (Chapter IX). These features include a long leader sequence, a variable amino terminus which is acetylated, a conserved hydrophobic core and a variable carboxy terminus. The leader peptide sequences are not as conserved as the pilin protein sequences. However, these leader peptides are all similar in size, ranging from 51 to 55 residues. All IncF pilin leader peptides contain a sequence which resembles a typical signal sequence (2) immediately upstream of the processing site. This part of the leader sequence consists of a positively charged amino acid followed by 18 hydrophobic residues, ending in a signal peptidase recognition sequence. Unpublished evidence (Chapter IX) indicates that the host-encoded signal peptidase may be capable of recognizing and cleaving at this site in the pED208 system. Other evidence suggests that the F system requires at least one other protein (traQ; 6) to achieve processing of the traA gene product by the signal peptidase. Further studies are required to establish whether the pED208 system encodes a traQ-like protein.

The only significant differences observed among F-like pili are located at the amino- and carboxy- termini (Chapter III, 4). However, these differences are sufficient to generate several pilus alleles (10). The hydrophobic core of these proteins is conserved, with few substitutions occurring. The less related pED208 pilin protein also varies considerably at the N- and C- termini when compared to F pilin (Chapter IX). Although most of pED208 pilin is hydrophobic, the central portion of these molecules is not strongly conserved (44%).

Armstrong et al. suggested that pED208 and F pili have similar conformations, based on circular dichroism and X-ray fiber diffraction studies (1). In Chapter III it was shown that ColB2 pilin could be assembled into a functional pilus by the tra products of the F plasmid, suggesting that these pilin proteins are similar in structure. Conservation of the hydrophobic core of these pilin proteins suggests that this region may be involved in a more "essential" role, such as protein folding and/or subunit-subunit interactions.

An acetylated amino terminus is a common feature of IncF pilin proteins (Chapter III, 3,5), with type I, II, III and IV pilins containing Ac-N-Ala amino termini, and type V containing an Ac-N-Thr residue. Conservation of this acetyl group implies that there is an analogous (or homologous) acetylating enzyme(s) involved in this modification. Laine et al. (6) have suggested that this protein may be the traG gene product in the F system. The pED208 system presumably encodes a similar enzyme.

TraL and traE are strongly conserved (nearly 100%) among the IncFI-IV plasmids, and share 44% homology with the pED208 system (Chapters III and IX, 4). Heteroduplex studies by Sharp et al. (8) indicate that the remainder of the region required for pilus assembly is conserved among IncFI-IV plasmids. This suggests that a common pilus-assembly mechanism is encoded by the IncFI-IV plasmids. Although this high level of homology does not extend to the pED208 system, its tra region does share numerous similarities with the F-like systems, including genetic ordering, gene function and homology. This indicates that pED208 has maintained an analogous pilus-assembly system, and the differences between these tra genes have probably occurred since these plasmids diverged during their evolution.

B. CONJUGAL DNA METABOLISM

Several alleles of genes involved in conjugal DNA transfer were presented in Chapters V, VII and IX. F-like traM genes encode proteins which are approximately 80-90% homologous to each other. The few differences in these proteins are sufficient to confer traM specificity, defining four traM alleles (10). However, traY alleles are much less conserved, varying significantly in size and amino acid composition. These differences also relect the various traY alleles defined by Willetts and Maule (10), presumably dictating traYZ endonuclease plasmid-specificity. The non-homologous sequences found in IncF traY and traM genes correlate well to the various oriT sequences. Some regions of IncF oriT sequences are completely conserved, others are highly homologous, and others are different. Comparison of these sequences allowed us to define regions which may be involved in protein-DNA interactions (Chapter VII) essential to conjugal DNA metabolism.

C. SURFACE EXCLUSION

Sequencing of the _traT_ gene from pED208 (Chapter VIII) provided us with the first evidence that the _tra_ region of pED208 plasmid is related to other IncF _tra_ regions. The _traT_ gene product of pED208 is 80% homologous to the equivalent products of F and R-100 plasmids (which differ by one amino acid). It would seem that the _traT_ gene is strongly conserved, even in less related systems. Willetts and Maule (9) suggested that the pilus tip may interact with TraTp, blocking conjugation. Until more IncF _traT_ product sequences are known, prediction of the regions of interaction between the _traT_ and _traA_ gene products is not possible. In some IncF plasmids the pilus and surface exclusion alleles do not match. These differences are probably attributable to _traS_ differences as Willetts and Maule did not differentiate between _traS_ and _traT_ functions when they defined surface exclusion alleles. Comparison of the pED208 _traS_ sequence with partial sequences of F and R-100 _traS_ revealed no homology between these genes. The region encoding this gene is different in IncF plasmids, based on heteroduplex analysis (8). It would seem that _traS_ is a "hypervariable" gene, differing extensively among the IncF plasmids. Whether these _traS_ genes all provide exclusion by similar mechanisms remains to be explored.

D. REGULATION OF CONJUGATION

The traJ/traYZ promoter loci also have different alleles which encode analogous functions (Chapter V; Frost, Finlay, and Paranchych, manuscript in preperation). The three IncFI-II traJ gene products cause increased transcription of their respective traYZ promoters, but share no homology. It is unclear why these related systems utilize different genes to achieve similar ends.

There is a strong correlation between finP and traJ alleles (10). However, finP allele specificity is determined at an RNA level (Chapter VI). The close linkage of finP and traJ allele groups is presumably due to the anti-sense RNA control of traJ by the finP product (and non-specific finO product), since finP and part of traJ must be encoded by the same DNA to produce complimentary RNA sequences.

E. EVOLUTION OF TRA ALLELE SPECIFICITY

It is difficult to establish an evolutionary order of IncF tra regions on the basis of tra allele homologies and differences. The extent of homology found between equivalent tra genes ranges from complete identity to significant differences. The tra alleles studied in this and other investigations can be ordered by comparison of their levels of gene and gene-product similarity. TraL and traE (and probably the other genes involved in pilus assembly) are nearly identical in IncFI-IV plasmids, with pED208 traL and traE sharing some homology with these genes. TraA, traT, traM and finP have diverged slightly more, with a few nucleotide or amino acid residues determining allele specificity. More divergent still are the traY genes. These genes contain homologous regions, but also encode several differences. Finally, traJ, the traYZ promoter and traS appear to be distinctly different, but, at least in the case of traJ/traYZ promoter, they encode analogous functions. Placement of oriT within this list is difficult since comparison of oriT sequences reveals regions of complete homology, as well significant variations. These similarities and differences reflect the extent of allele homology of the proteins which interact with this region of DNA.

Comparison of related sequences from IncF tra genes indicates that the pED208 tra region diverged from IncFI-IV tra regions before IncFI-IV allele specificity developed. However, within the IncFI-IV tra genes, one cannot establish an order of divergence. For example, the first part of types III and V finP sequences are more closely related than the equivalent sequences in the type IV allele (Chapter VI). However, the proximal end of types III and IV finP genes are homologous, and the type V sequence differs. Another example of these varying homologies is seen with the oriT sequences (Chapter VII).

It is possible that these plasmids may have exchanged some of their tra genes through recombination mechanisms. Since these plasmids are continually entering and exiting host bacteria, ample opportunity exists for two compatible plasmids to be present in a cell allowing recombination to occur between tra regions. However, this mechanism does not explain how the tra allele differences arose, suggesting that other mechanisms for allele diversification exist.

These differences may have arisen by divergence of "non-functional" domains of a protein, while the "essential" parts of a protein remained conserved. Support for this argument is provided by the traA allele group in which the variable N-terminus appears to be nonfunctional, although the conserved N-terminal acetyl group may be essential. Contradicting this argument are the traJ sequences which differ significantly among IncF plasmids. TraJ gene products share only small amounts of homology, (approximately 20%), indicating that these genes are an evolutionary "hotspot", undergoing many changes, while other tra regions remain virtually unaltered.

Insertion elements may also play a role in tra operon evolution. IS2 elements are present in pED208 traY (Chapter IX), R-100 traE (unpublished) and between R-100 traS and traT (7). These elements affect regulation more than protein function, significantly altering levels of various tra proteins.

In conclusion, the generation of allele specificity within the genes required for conjugative DNA transfer must have occurred by several mechanisms as these plasmids evolved. Examination of some of these differences explains the heteroduplex results reported by Sharp *et al*. (8) and the allele groupings defined by Willetts and Maule (10). More significantly, this study has provided us with a clearer understanding of the complex biological processes involved in conjugation.

F. BIBLIOGRAPHY

1. **Armstrong, G.D., L.S. Frost, P.A. Sastry, and W. Paranchych**. 1980. Comparative biochemical studies on F and EDP208 conjugative pili. J. Bact. **141**:333-341.

2. **Benson, S.A., M.N. Hall, and T.J. Silhavy**. 1985. Genetic analysis of protein export in _Escherichia_ _coli_ K-12. Ann. Rev. Bioch. **54**:101-134.

3. **Frost, L.S., G.D. Armstrong, B.B. Finlay, B.F.P. Edwards, and W. Paranchych**. 1983. N-terminal amino acid sequencing of EDP208 conjugative pili. J. Bacteriol. **153**:950-954.

4. **Frost, L.S., B.B. Finlay, A. Opgenorth, W. Paranchych and J. Lee**. 1985. Characterization and sequence analysis of pilin from IncF plasmids. J. Bacteriol. **164**:1238-1247.

5. **Frost, L.S., W. Paranchych, and N.S. Willetts**. 1984. DNA sequence of the F _traALE_ region that incudes the gene for F pilin. J. Bacteriol. **160**:395-401.

6. **Laine, S., D. Moore, P. Kathir, and K. Ippen-Ihler**. 1985. Genes and gene products involved in the synthesis of F-pili, p. 535-553. _In_ D. Helinski, S. Cohen, D. Clewell, D. Jackson, and A. Hollaender (eds.), Plasmids in Bacteria. Plenum Publishing Company, New York.

7. **Ogata, R., C. Winters, and R.P. Levine**. 1982. Nucleotide sequence analysis of the complement resistance gene from plasmid R100. J. Bacteriol. **151**:819-827.

8. **Sharp, P.A., S.N. Cohen, and N. Davidson.** 1973. Electron microscope heteroduplex studies of sequence relations among plasmids of E. coli II. Studies of drug resistant (R) factors and F factors. J. Mol. Biol. **75**:235-255.

9. **Willetts, N., and J. Maule.** 1974. Interactions between the surface exclusion systems of some F-like plasmids. Genet. Res. **24**:81-89.

10. **Willetts, N.S., and J. Maule.** 1986. Specificities of incF plasmid conjugation genes. Genet. Res. **47**:1-11.

APPENDIX

PUBLICATIONS

1. Frost, L., Armstrong, G., Finlay, B., Edwards, B., Paranchych, W. 1983. N-terminal amino acid sequencing of EDP208 conjugative pili. J. Bacteriol. **153**: 950-954.

2. Finlay, B., Paranchych, W., Falkow, S. 1983. Characterization of conjugative plasmid EDP208. J. Bacteriol. **156**: 230-235.

3. Finlay, B., Frost, L., Paranchych, W. 1984. Localization, cloning and sequencing of the conjugative plasmid ColB2 pilin gene. J. Bacteriol. **160**:402-407.

4. Pasloske, B., Finlay, B., Paranchych, W. 1985. Cloning and sequencing of the *Pseudomonas aeruginosa* PAK pilin gene. FEBS Letters **183**:408-412.

5. Finlay, B., Frost, L., Paranchych, W., Parker, J., Hodges, R. 1985. The major antigenic determinants of F and ColB2 pili. J. Bacteriol. **163**:331-335.

6. Finlay, B., Pasloske, B., Paranchych, W. 1986. Expression of *Pseudomonas aeruginosa* PAK pilin gene in *Escherichia coli*. J. Bacteriol. **165**:625-630.

7. Sastry, P., Finlay, B., Pasloske, B., Paranchych, W., Pearlstone, J., Smillie, L. 1985. Comparative studies of the amino acid and nucleotide sequences of pilin derived from *Pseudomonas aeruginosa* strains PAK and PAO. J. Bacteriol. **164**: 571-577.

8. Frost, L., Finlay, B., Opgenorth, A., Paranchych, W., Lee, J. 1985. Characterization and sequence analysis of pilin from IncF plasmids. J. Bacteriol. **164**: 1238-1247.

9. Finlay, B., Frost, L., Paranchych, W. 1986. Nucleotide sequence of the R1-19 plasmid transfer genes *traM*, *finP*, *traJ*, *traY* and *traYZ* promoter. J. Bacteriol. May issue.

10. Finlay, B., Frost, L., Paranchych, W., Willetts, N.S. 1986. Nucleotide sequence of five IncF *finP* alleles. J. Bacteriol. August Issue.

11. Finlay, B., Paranchych, W. 1986. Nucleotide sequence of the surface exclusion genes *traS* and *traT* from the IncFV plasmid pED208. J. Bacteriol. June issue.

12. Finlay, B., Frost, L., Paranchych, W. Studies on the origin of transfer of IncF plasmids: Nucleotide sequences of the typeII *oriT*/*traM*/*traY* allele from ColB4-K98 and the typeIV *traY* allele from R-100-1. J. Bacteriol. (submitted).

13. Lobe, C., Finlay, B., Paranchych, W., Paetkau, V., Bleakley, C. 1986. Two cytotoxic T lymphocyte-specific genes encode unique serine proteases. Science (accepted).

14. Frost, L., Worobec, E., Pieroni, P., Armstrong, G., Hodges, R., Parker, J., Finlay, B., Paranchych, W. 1986. The location of the antigenic determinants on the surface of conjugative F-like pili. J. Bacteriol. (submitted).

15. Finlay, B., Frost, L., Paranchych, W. Nucleotide sequence of the *traYALE* region from the IncFV plasmid pED208. J. Bacteriol. (submitted).

16. Frost, L., Finlay, B., Paranchych, W. Nucleotide sequence of *traJ*, *traE* and the *traYZ* promoter of the R100-1 plasmid. (in preparation).

17. Paranchych, W., Finlay, B., and Frost, L. 1986. Studies on the regulation of IncF plasmid operon expression. *In* Banbury Report on Evolution and Environmental Spread of Antibiotic resistance genes. In press.